

Untersuchung von mono- und multivariaten Methoden zur Signalloptimierung am Beispiel des Zerfalls $B^\pm \rightarrow J/\psi K^\pm$ im ATLAS-Detektor

zur Erlangung des akademischen Grades
Bachelor of Science



Universität Siegen
Fakultät IV

Autor: David Fritz
E-Mail: david.fritz@student.uni-siegen.de
Matrikelnummer: 1517260

Version vom: 13. Oktober 2022

Betreuer: Prof. Dr. Markus Cristinziani

Inhaltsverzeichnis

1	Einleitung	3
2	Das ATLAS-Experiment	4
2.1	Der Large Hadron Collider	4
2.2	Der Aufbau des ATLAS-Detektors	5
2.2.1	Der innere Detektor	6
2.2.2	Das Kalorimeter	8
2.2.3	Das Myonen-Spektrometer	9
3	Theoretische Grundlagen und statistische Methoden	10
3.1	Das Standardmodell	10
3.1.1	Der Aufbau des Standardmodells	10
3.1.2	Wechselwirkungen	12
3.1.3	Der Higgs-Mechanismus	15
3.2	Die Fisher-Diskriminante	15
3.2.1	Ausgangsproblem	15
3.2.2	Mathematisches Vorgehen	16
3.3	Boosted Decision Trees	17
3.4	Grenzwertoptimierungskurve	18
4	Der untersuchte Zerfall	19
4.1	Die untersuchten Größen	20
5	Optimierung	22
5.1	Verwendete Datensätze	22
5.2	Schnitt auf einen Parameter	26
5.2.1	Training auf Monte-Carlo-Daten	26
5.2.2	Anwendung auf reale Messdaten	27
5.3	Die Fisher-Diskriminante	30
5.3.1	Training auf Monte-Carlo-Daten	30
5.3.2	Eigene Implementierung der Fisher Diskriminante	30
5.3.3	TMVA: Fisher Diskriminante	32
5.3.4	Anwendung auf reale Messdaten	34
5.4	Boosted Decision Trees	37
5.4.1	Training auf Monte-Carlo-Daten	37
5.4.2	Anwendung auf reale Messdaten	39
6	Vergleich der Methoden	40
7	Zusammenfassung und Ausblick	42
8	Anhang	44
9	Literaturverzeichnis	49
10	Abkürzungsverzeichnis	50

1 Einleitung

Die Teilchenbeschleuniger am Europäischen Teilchenforschungszentrum CERN, das 1954 nahe Genf gegründet wurde [1], und viele weitere Beschleuniger weltweit haben in den letzten Jahrzehnten zwei große gemeinsame Aufgaben verfolgt: Das Überprüfen der theoretischen Vorhersagen des Standardmodells der Teilchenphysik und die unabhängige Suche nach neuen Phänomenen. Nach unzähligen Kollisionen und Messungen konnte das Standardmodell in vielen seiner Vorhersagen bestätigt werden. Daher gilt es als eines der vielversprechendsten theoretischen Modelle der Teilchenphysik.

Es konnten bis heute die durch Murray Gell-Mann 1964 postulierten Up- und Down-Quarks [2], sowie zwei weitere Familien von Quark-Dubletts identifiziert werden, die sich nur in ihrer wachsenden Masse unterscheiden. Das bisher schwerste und letzte Top-Quark wurde erst 1995 entdeckt. Zumindest wird nicht von der Existenz einer schwereren Familie ausgegangen.

Heute befinden sich am LHC (Large Hadron Collider), welcher vorrangig Protonen zur Kollision bringt, vier große Experimente: CMS (Compact Muon Solenoid), ATLAS (A Toroidal LHC ApparatuS), LHCb (Large Hadron Collider beauty) und ALICE (A Large Ion Collider Experiment) [3] sowie weitere kleinere Experimente. ATLAS, genauso wie CMS, LHCb und ALICE, spielt eine entscheidende Rolle in der Suche nach neuer Physik, neuer Teilchen und bisher unentdeckter Phänomene. Es handelt sich um einen hoch sensiblen Detektor, welcher die Teilchen von mehr als einer Milliarde Kollisionen pro Sekunde detektiert. Daher ist die computergestützte Filterung nach der Rekonstruktion von Spuren und Wechselwirkungspunkten von tatsächlichen Signalen und Untergrund der Messung von großer Bedeutung. Sie hat bspw. erst die Entdeckung des Higgs-Bosons 2012 in ATLAS und CMS möglich gemacht [4] [5].

Ziel dieser Arbeit soll der Vergleich unterschiedlicher Methoden der Trennung von Signal und Untergrund sein, in dem ihre Effizienzen und Untergrundunterdrückungsraten analysiert werden. Dabei kommen drei zunehmend komplexere Methoden zum Einsatz. Die Erste bedient sich Schnitte auf einzelne Variablen. Die Zweite, die lineare Diskriminanzfunktion nach Fisher [6], verwendet Zusammenhänge zwischen den Parametern für eine optimale Trennung. Die letzte Methode nutzt Machine Learning (BDT) [7] um eine komplexe und dadurch erwartungsgemäß effektivere Trennung von Signal und Untergrund zu ermöglichen.

Diese Methoden werden zunächst auf simulierten Daten trainiert, ehe sie im Anschluss auf Messdaten des ATLAS-Detektors Anwendung finden. Für die Implementierung dieser Methoden wird ROOT verwendet, einem eigens für die Analyse und Darstellung von Messdaten entwickeltes Toolkit auf Basis von C++ [8].

2 Das ATLAS-Experiment

ATLAS [9] ist weltweit der größte Detektor für Teilchenkollisionen. Er enthält mehr als 100 Millionen einzelne Signalkanäle, um Teilchenkollisionen zu registrieren [9]. Er wurde 2009 fertiggestellt und zeichnete im gleichen Jahr erste Kollisionen auf. Er befindet sich in einer Tiefe von rund 100 m im Tunnel des LHC, nahe der Stadt Genf (Abb. 2).

Er dient der Untersuchung eines breiten Spektrums an Forschungsthemen, darunter die Suche nach exotischen Teilchen, Dunkler Materie oder Anzeichen für Abweichungen vom Standardmodell. Im ATLAS-Detektor wurde 2012 erstmals ein Teilchen nachgewiesen, dessen Eigenschaften mit dem Higgs-Boson übereinstimmen [5]. Neben ATLAS wird CMS ebenso für diese Forschungsbereiche eingesetzt [4]. Beide befinden sich symmetrisch an zwei Kollisionspunkten im Tunnel, sodass es zeitgleich zu Kollisionen kommt. CMS wird ebenfalls für diese Forschungsschwerpunkte verwendet, um Ergebnisse unabhängig zu überprüfen und zu bestätigen. Das CMS Experiment verwendet komplementäre Technologien zu ATLAS.

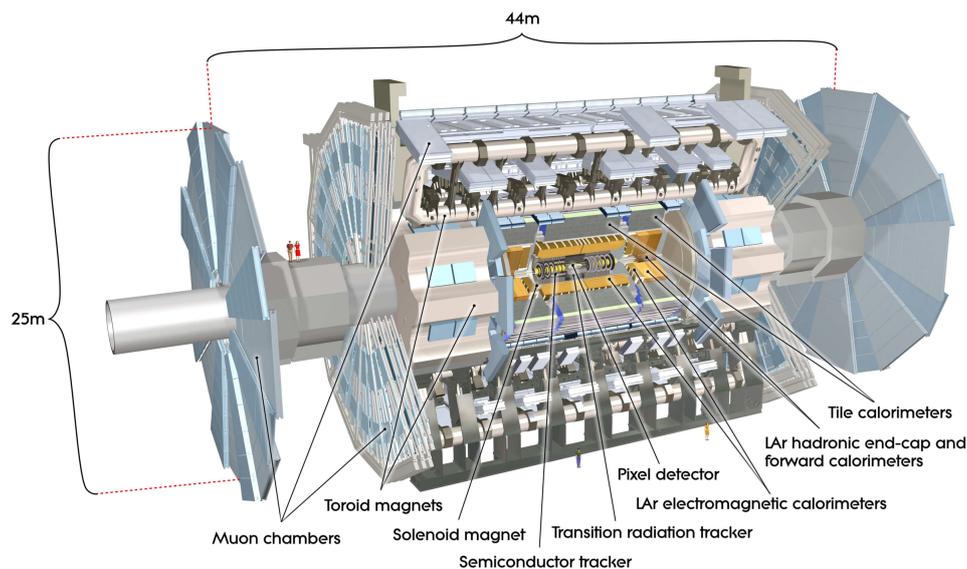


Abbildung 1: Schematischer Aufbau des ATLAS-Detektors mit dem inneren Solenoid-Spulensystem im Zentrum und den darum schalenförmig angeordneten Kalorimetern [9].

2.1 Der Large Hadron Collider

Der Large Hadron Collider ist ein rund 26,65 km langer Beschleunigerring [3], der sich in einer durchschnittlichen Tiefe von rund 100 m unter dem Schweizer Kanton Genf und dem französischen Grenzgebiet befindet (Abb. 2). Mittels Linearbeschleunigern (LINAC) werden Protonen, die zuvor in der Duoplasmatron Protonenquelle in Bündeln von rund $1,15 \cdot 10^{11}$ Teilchen erzeugt wurden, vorbeschleunigt, bevor sie im Anschluss über mehrere Zwischenstufen in Protonen Synchrotrons, wie PSB, PS und SPS auf 450 GeV beschleunigt werden [3]. Nach Injektion in den LHC werden die Protonen weiter gebündelt und auf eine Schwerpunktsenergie von 6,8 TeV gebracht, dies entspricht einer totalen Strahl-

energie von 13,6 TeV in den gegenläufigen Protonenstrahlen. Diese Schwerpunktsenergien wurden erstmals im April 2022, während Run 3 des LHC nach einer 3 jährigen Ausbauphase erreicht. Während Run 2 erreichten die Protonen eine Energie von 6,5 TeV. Damit werden Luminositäten von $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ erreicht.

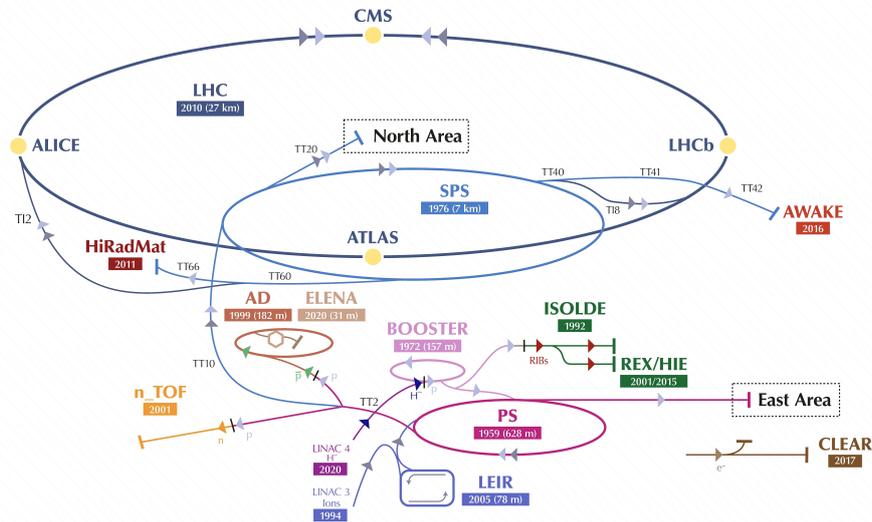


Abbildung 2: Der LHC: eingezeichnet sind die Positionen der vier großen Experimente darunter ATLAS und CMS, die sich gegenüber befinden, sowie weitere kleine Experimente und die Vorbeschleuniger [10].

Wenn die Protonen-Haufen, von denen maximal 2808 im Beschleunigerring Platz finden, sich im LHC in zwei Partikelstrahlen bewegen, besitzen sie eine Geschwindigkeit von 99,9999991 Prozent der Lichtgeschwindigkeit mit einer kinetischen Energie von 6,8 TeV pro Proton [3]. In zwei parallel verlaufenden Röhren laufen diese Protonenbündel gegenläufig durch die Beschleunigerstrecken, und die Fokussier- und Umlenkmagneten ehe sie an den Wechselwirkungspunkten in den vier Detektoren zur Kollision gebracht werden.

2.2 Der Aufbau des ATLAS-Detektors

Der ATLAS-Detektor [9] besitzt einen Umfang von ca. 25 m, ist ca. 46 m lang und 7000 t schwer. Er ist zylindrisch um die Strahlachse, welche der z-Achse entspricht, aufgebaut. Die gegenläufigen, im LHC beschleunigten, Protonenstrahlen kollidieren in der Mitte des ATLAS-Detektors und erzeugen Leptonen und Jets neuer Partikel, die vom primären Wechselwirkungspunkt in alle Richtungen fliegen und zum Teil an sekundären Wechselwirkungspunkten weiter zerfallen. Diese Leptonen und Jets werden von sechs unterschiedlichen Subsystemen vermessen.

Das Zentrum des Detektors bildet der innere Detektor (Abb. 3). Dieser Teil des Detektors ist in unmittelbarer Nähe der Kollisionen und zeichnet die Spurpunkte geladener Teilchen auf, auch solche, die nach der Kollision nach einigen Zentimetern zerfallen. Um den Pixeldetektor herum befinden sich der Silizium-Streifen-Detektor (SCT) und der Transition Radiation Tracker (TRT) [9]. Vor dem elektromagnetischen (ECAL) und dem hadronischen Kalorimeter (HCAL) befindet sich noch die Solenoid-Spule, die ein möglichst homogenes Magnetfeld im inneren Detektor erzeugt (Abb. 6). Das ECAL befindet

sich näher an der Strahlachse, da die Eindringtiefe der Elektronen geringer ist als die der Hadronen, welche zum größten Teil ihre Energie im HCAL deponieren. Aufgrund der mehrlagigen Anordnung der Kalorimeterzellen können die kinetischen Energien rekonstruiert werden. Das durch solenoid und toroidal positionierte Spulensysteme entstehende Magnetfeld lenkt Teilchen unterschiedlicher Impulse entsprechend ihrer Geschwindigkeit und Ladung durch die Lorentzkraft ab. In der Folge kann durch die Krümmung der Trajektorie der Betrag des Impulses, die kinetische Energie und der Abstand des zweiten Wechselwirkungspunktes zum Ersten, ein Teilchen genau charakterisiert werden. Den äußeren Mantel des Detektors bilden die Myonkammern, da Myonen nur wenig Ladung in den genannten Detektoren verlieren. Diese Myonendetektoren sind weit vom Streuzentrum entfernt angebracht (Abb. 6). Die Myonenkammern sind vor allem innerhalb der Toroidspulen angebracht.

Es entsteht eine erhebliche Datenmenge während der Kollisionen im ATLAS-Detektor. Um eine erste Selektion der Daten vorzunehmen, werden die Messdaten an mehreren Stellen, unter anderem in den Detektormodulen und den Auslesekanälen gepuffert, bevor das Triggersystem über potenziell interessante Kollisionen entscheidet. Komplexe Programme und Algorithmen werden verwendet, um die aufgezeichneten Ereignisse zu analysieren und weiter zu verarbeiten.

2.2.1 Der innere Detektor

Im Zentrum des Detektors befindet sich der innere Detektor, welcher in drei Subsysteme aufgeteilt ist (Abb. 3). Er befindet sich in unmittelbarer Nähe zum Wechselwirkungspunkt, um eine präzise Rekonstruktion der Trajektorien geladener Teilchen zu ermöglichen. Er misst circa 7 m in Länge und hat einen Durchmesser von circa 2 m [9]. Ein Solenoidmagnet erzeugt im inneren Detektor, den er umschließt, ein homogenes Magnetfeld von 2 T, das eine Krümmung der Teilchenspuren verursacht. Die drei Subsysteme sind konzentrisch um die Strahlachse angeordnet (Abb. 3). Der innere Detektor besteht aus dem Pixeldetektor, dem Silizium-Streifen-Detektor und dem Transition Radiation Tracker [3].

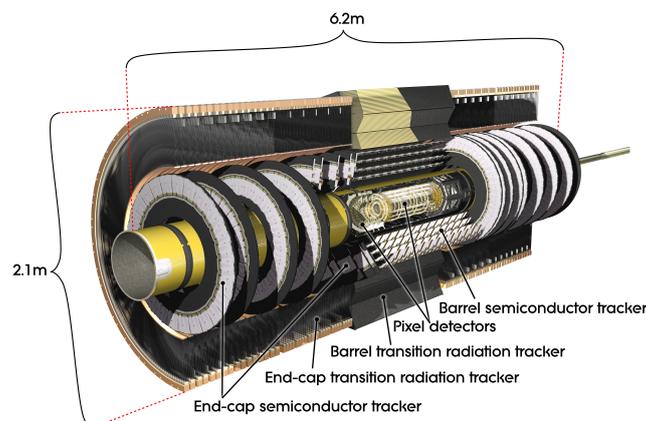


Abbildung 3: Schematischer Aufbau des inneren Detektors [9].

1. Der Pixeldetektor: Dieser Detektor setzt sich aus 2200 Auslesemodulen zusammen, die jeweils 61440 Pixel tragen, und in vier Lagen konzentrisch um die Strahlachse angeordnet sind [3]. Die ca. 140 Millionen Pixel haben eine typische Größe von $50 \mu\text{m} \times 400 \mu\text{m}$

[9]. Der Pixeldetektor befindet sich am nächsten zum Wechselwirkungspunkt (Abb. 4) und wird an der Vorder- und Endkappe jeweils durch drei weitere Scheiben von Pixeln abgeschlossen. Die Module des Pixeldetektors bestehen aus Silizium-Halbleiterdetektoren, die aufgrund ihrer geringen Dicke perfekt für die Anwendung im inneren des Detektors geeignet sind. Sie ermöglichen eine präzise Auflösung der Teilchentrajektorien, was gerade bei dem geringen Abstand vom Streuzentrum essenzielle Bedeutung zukommt. Besonders Mesonen und Baryonen haben aufgrund ihrer Masse nur eine geringe Lebensdauer, die durch Zeitdilatation eine Zerfallslänge von einigen $100 \mu\text{m}$ haben. Auch die kurzen Zeitabstände zwischen den Proton-Proton-Kollisionen von $2,5 \text{ ns}$, erschweren das Bestimmen dieser Teilchen, da es zur Anhäufung an Messdaten in den Puffern kommt, ehe diese durch das Triggersystem abgerufen werden.

2. Der Silizium-Streifen-Detektor (SCT): Dieses zweite System des inneren Detektors besteht aus Siliziumdetektoren und ist in jeweils 9 Scheiben in Vorwärts- und Rückwärts-Richtung sowie in vier Lagen um den Pixeldetektor aufgeteilt (Abb. 4). Er besitzt eine Auflösung von $16 \mu\text{m}$ radial zum Strahl [9]. Die Auflösung entlang der Strahlachse beträgt $580 \mu\text{m}$ [9]. Die einzelnen Streifenlagen sind gegeneinander um $40 \mu\text{m}$ verschoben. Damit erreicht der SCT eine Genauigkeit von $30 \mu\text{m}$ [9] bei den Spurkoordinaten. Er besitzt 4088 einzelne Auslesemodule mit jeweils 768 einzelnen Kanälen. Er hat ein einer Diode verwandtes Funktionsprinzip und wird durch die Ionisierung des Halbleitermaterials ausgelöst.

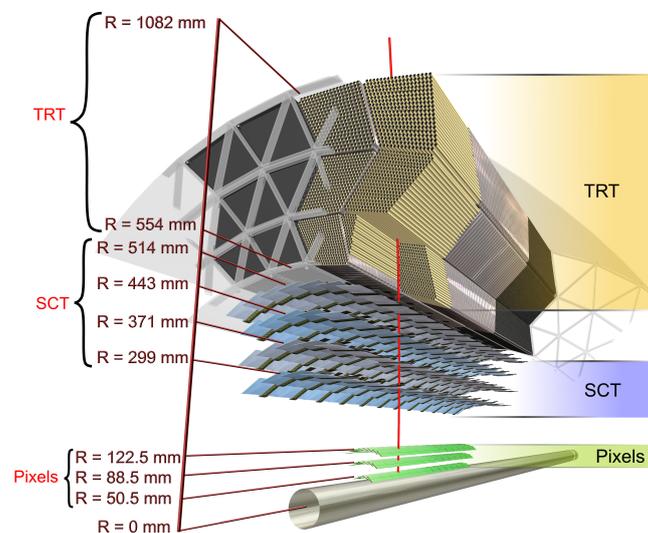


Abbildung 4: Der innere Detektor im Querschnitt. Im direkten Umfeld der Strahlachse befindet sich der Pixeldetektor (mittlerweile wurde eine vierte Lage hinzugefügt). Danach folgen zunächst der SCT und dann der TRT [9].

3. Der Transition Radiation Tracker (TRT): Der TRT ist ebenfalls konzentrisch um die Strahlachse montiert und besteht aus 50000 etwa 4 mm kleinen Driftröhrchen. Es bildet den größten Teil des inneren Detektors (Abb. 4). Diese Röhrchen geben aufgrund von Relaxation angeregter Atome des Materials auf Polyethylen-Basis, verursacht durch

Elektronen, ein Signal aus. Sie sind mit einem Xenongasgemisch befüllt, das mit den vorher emittierten Röntgenphotonen wechselwirkt [3]. Das Funktionsprinzip ähnelt dem einer Driftkammer, in der ebenfalls Teilchenspuren durch Photonemission sichtbar gemacht werden. Die Auflösung der Spurkoordinaten im TRT beträgt ungefähr $150 \mu\text{m}$ [9].

2.2.2 Das Kalorimeter

Das Kalorimeter ummantelt den inneren Detektor sowie den Solenoidmagneten vollkommen und besteht aus zwei Detektorsystemen. Das elektromagnetische um den Solenoidmagneten und das hadronische Kalorimeter dahinter dienen der Messung der durch Elektronen und Photonen im ECAL und hadronische Teilchen im HCAL deponierten Energie. Dabei messen sie lediglich ein vorher fein kalibriertes Signal, dessen Stärke proportional zur deponierten Energie der Teilchen ist.

Das elektromagnetische Kalorimeter (ECAL) befindet sich am Nächsten am inneren Detektor (Abb. 5). Es misst die Energien der produzierten Elektronen und Photonen. Zur Signalerfassung wird als Medium flüssiges Argon in den Zellen verwendet, während die Absorberschichten aus Kupfer, Blei und Wolfram bestehen. Eine typische der insgesamt 190.000 Zellen hat eine Querschnittsfläche von rund 6 cm^2 [9]. Die Zellen besitzen eine einzigartige Akkordeon-ähnliche Geometrie, bei der die Absorberplatten verdreht zur radial orthogonalen Orientierung positioniert werden, um eine gleichmäßige Abdeckung unabhängig der Einfallsrichtung des Teilchens zu gewährleisten [9].

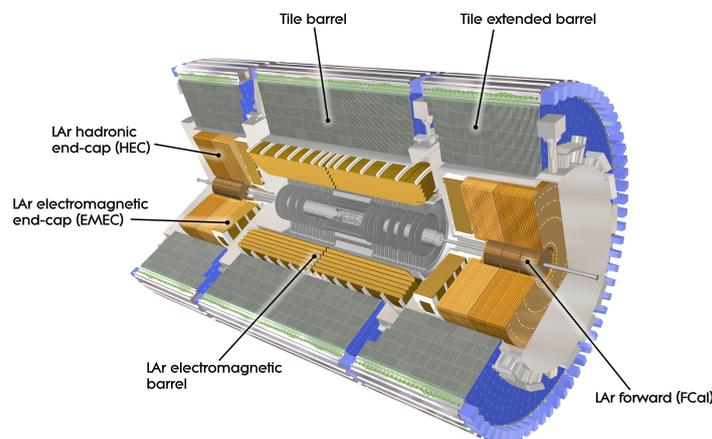


Abbildung 5: Die goldenen inneren Bereiche gehören zum ECAL, während die grauen äußeren Bereiche das HCAL zeigen [9].

Das hadronische Kalorimeter (HCAL) umschließt das elektromagnetische Kalorimeter und ist dafür verantwortlich, die Energie der Hadronen zu messen, die den größten Teil ihrer Energie dort deponieren (Abb. 5). Es befindet sich weiter außerhalb und hat eine größere Granularität als das ECAL, da nur ein geringer Teil der Energie der Hadronen im ECAL deponiert wird. Anders als das elektromagnetische Kalorimeter, setzt sich das hadronische Kalorimeter aus Eisenabsorbieren und Szintillatoren aus Polyethylen zusammen [9]. Die Vorwärtsabsorber und die Endkappen des Kalorimeters bedienen sich allerdings der gleichen Technologie wie das ECAL. Aufgrund der hohen kinetischen Energien ist das ha-

dronische Kalorimeter räumlich wesentlich größer als das elektromagnetische Kalorimeter und besitzt größere Auslesezellen.

2.2.3 Das Myonen-Spektrometer

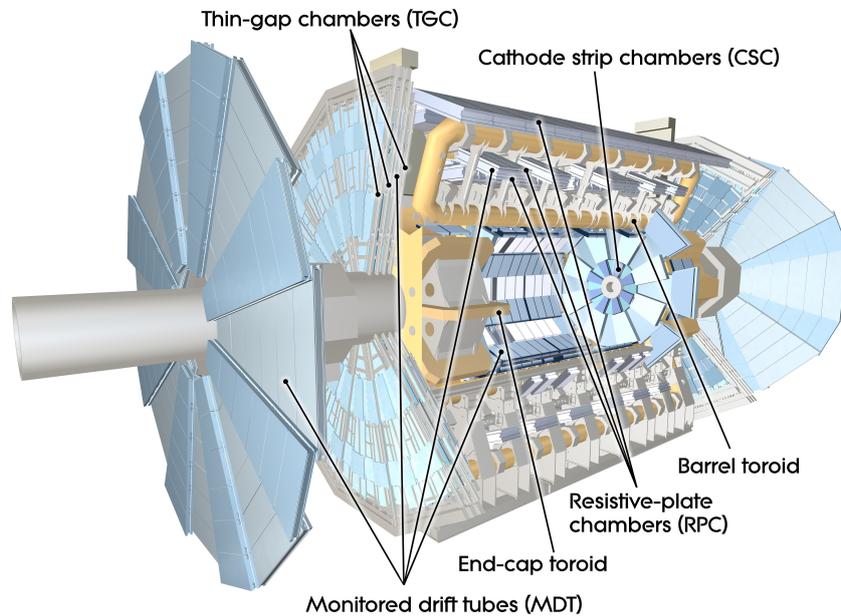


Abbildung 6: Aufbau des Myonenspektrometers und des toroidalen Spulensystems im ATLAS-Detektor [9].

Die äußerste Schale des ATLAS-Detektors bildet das Myonen Spektrometer, das sich aus drei verschiedenen Myonendetektoren zusammensetzt. Es dient der Erfassung von Myonen und der Bestimmung ihres Impulses (Abb. 6) [9].

Die Myondetektoren sind in der äußersten Schale angebracht, da Myonen aufgrund ihrer großen Masse von 105,65 MeV [11] eine verhältnismäßig große Zerfallslänge haben und nur wenig Energie in den inneren Teilen des Detektors deponieren. Die Myonendetektoren befinden sich in einem Magnetfeld, welches durch ein toroidales Spulensystem erzeugt wird und die Messung des Impulses erst ermöglicht (Abb. 6).

Die einzelnen MDT-Module (Monitored Drift Tube) haben eine Länge zwischen 1 m und 6 m [9]. Die Driftröhrchen sind in drei Lagen angeordnet, die insgesamt mit eine Million Auslesezellen eine transversale Auflösung von $50 \mu\text{m}$ erzielen [9], welche eine besonders genaue Rekonstruktion der Teilchenspuren ermöglicht. An den Enden des Detektors befinden sich zudem drei Scheiben aus MDT-Modulen (Abb. 6). Im Unterschied zu den anderen zwei Detektorsystemen besitzt das Myonspektrometer ein eigenes Triggersystem, das in Zusammenarbeit mit dem Vortrigger die Selektion der Kollisionsdaten vornimmt.

3 Theoretische Grundlagen und statistische Methoden

Eine der wichtigsten Grundlagen für die experimentelle Teilchenphysik, ist eine Theorie, die präzise Vorhersagen treffen kann, für deren Überprüfbarkeit Detektoren wie ATLAS gebaut wurden. Diese, der Teilchenphysik bis heute als eine der erfolgreichsten Konzepte zugrundeliegende Theorie, ist das Standardmodell der Teilchenphysik. Dennoch regen unerwartete Ergebnisse in den Experimenten auch die Weiterentwicklung der Theorie und das Aufstellen neuer Hypothesen an. Dieser Abschnitt soll einen kurzen Überblick über die wichtigsten Annahmen und Bestandteile des Standardmodells geben sowie einen Einblick in die bisher bekannten Materieteilchen und Wechselwirkungen.

Aufgrund der großen Datenmengen müssen die Daten selektiert werden, um die vorhergesagten Signale vom Untergrund trennen zu können und die Theorie überprüfen zu können. Eine der einfachsten mehrdimensionalen Methoden statistischer Separation ist die lineare Diskriminanzfunktion (engl. "Fisher discriminant"), die an dieser Stelle kurz hergeleitet und begründet werden soll. Um das Vergleichen von Methoden zu veranschaulichen, soll zudem ein ROC-Graph (Receiver Operator Characteristic) verwendet werden.

3.1 Das Standardmodell

Das Standardmodell ist die beste Beschreibung der Materie und der Wechselwirkungen bis heute und konnte schon unzählige korrekte Vorhersagen treffen. Es vereint zwei wichtige Theorien der Materie und Wechselwirkungen: Die Quantenchromodynamik (QCD) und die elektroschwache Kraft [12]. Durch diesen Zusammenschluss ist es möglich drei der vier grundlegenden Kräfte der Physik, die starke Wechselwirkung und die in der elektroschwachen Wechselwirkung vereinte elektromagnetische Wechselwirkung und schwache Wechselwirkung, theoretisch vollständig zu beschreiben. Es bildet zudem einen Rahmen für weitere Entwicklungen und Verbesserungen. Die vier grundlegenden Kräfte sind die makroskopisch wirkenden Kräfte: Die elektromagnetische Kraft, die durch Ladungen verursacht wird, sowie die Gravitationskraft, welche sich durch die Anziehung von Massen äußert. Die zwei übrigen Kräfte sind die schwache und starke Kraft, deren Wirkungsradien in der Größenordnung 10^{-10} m liegen [12]. Das Standardmodell vereint die elektromagnetische und schwache Kraft in einer Wechselwirkung und beschreibt diese sowie die starke Kraft. Die Gravitationskraft kann nicht mit dem Standardmodell in dessen jetziger Form in Einklang gebracht werden. Zur Vereinigung aller Kräfte in einer Weltformel gibt es Bestrebungen, deren bekannteste Vertreter die String-Theorie und Quantengravitation sind.

3.1.1 Der Aufbau des Standardmodells

Das Standardmodell umfasst 17 bisher bekannte Elementarteilchen und 13 Antiteilchen (Abb. 7), darunter befinden sich drei Generationen von jeweils vier materiebildenden Elementarteilchen, den zwei Quarks und den zwei Leptonen. Die Eichbosonen, die W^\pm - und Z-Bosonen, sind Austauscheteilchen der schwachen Wechselwirkung [12]. Das Austauscheteilchen der elektromagnetischen Wechselwirkung ist das Photon, während das Gluon das Austauscheteilchen der starken Wechselwirkung ist. Alle Eichbosonen zählen zu den Vektorbosonen. Das übrig bleibende Teilchen ist das skalare Higgs-Boson. Das Higgs-Feld verleiht allen Teilchen des Standardmodells, bis auf das Photon und die Neutrinos, ihre Masse.

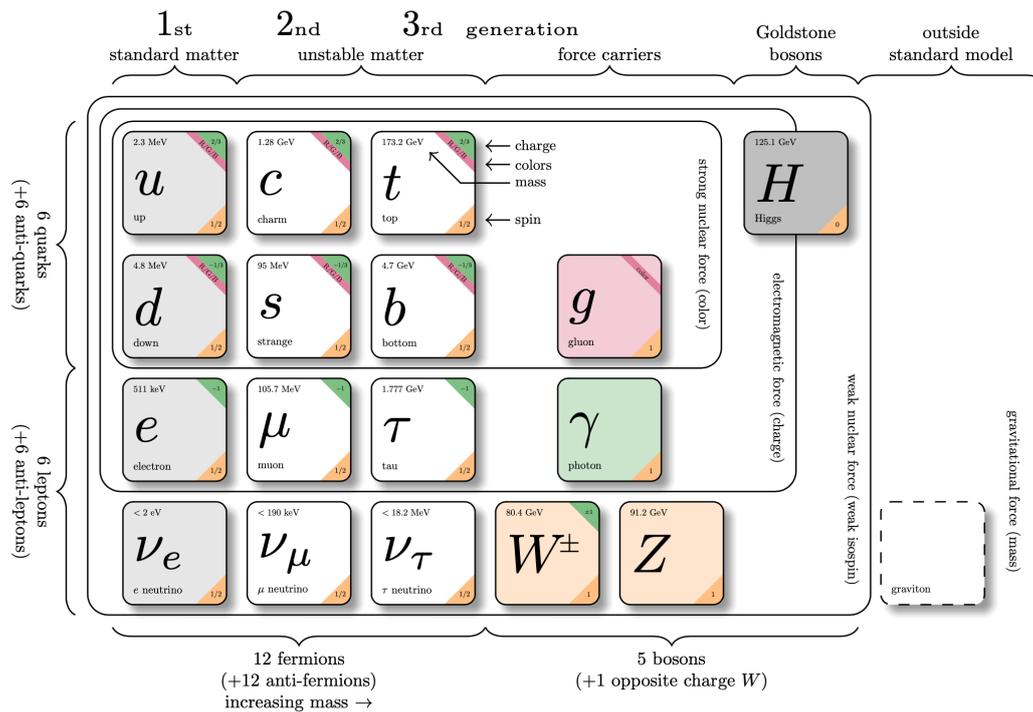


Abbildung 7: Das Standardmodell der Teilchenphysik mit allen bisher 17 charakterisierten Elementarteilchen und ihren Massen sowie die zugehörigen 13 Antiteilchen [13].

Die 12 Fermionen, entsprechend Teilchen mit einem halbzahligem Spin, bilden 3 Generationen von quasi-identischen Teilchen, die sich allein in ihrer Masse unterscheiden. Die erste Generation, aus der die uns umgebende Materie besteht, wird von zwei Quarks, dem Up- und Down-Quark, und zwei sogenannten Leptonen, dem Elektron und dem Elektron-Neutrino gebildet. Das Up-Quark besitzt eine Ladung von $\frac{2}{3}e$, während das Down-Quark eine Ladung von $-\frac{1}{3}e$ besitzt (Abb. 7). Die beiden Quarks der ersten Generation werden als Dublett dargestellt (Gl. 1). Die Nukleonen, die den Atomkern aller Elemente bilden, werden jeweils aus einer Kombination von drei Valenz-Quarks gebildet. Die Protonen bestehen aus zwei Up-Quarks und einem Down-Quark. Das Neutron setzt sich aus zwei Down-Quarks und einem Up-Quark zusammen, sodass seine Ladung neutral ist [12]. Zu allen Teilchen der materiebildenden Fermionen existieren auch entsprechende Anti-Teilchen, welche physikalisch bis auf eine entgegengesetzte Ladung die gleichen Eigenschaften haben und somit Antimaterie bilden.

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix} \quad (1)$$

Neben der ersten Generation existieren noch zwei weitaus schwerere Kopien, die allerdings nur in Beschleunigern oder anderen hochenergetischen Prozessen erzeugt werden. Die zweite Generation besteht aus dem Charm-Quark, das die gleichen Eigenschaften wie das Up-Quark besitzt, dem Strange-Quark, das die gleichen Eigenschaften hat, wie das Down-Quark, sowie dem Myon und dem Myon-Neutrino, die sich durch die jeweiligen Massen vom Elektron und Elektron-Neutrino unterscheiden [12]. Das Myon ist das einzige

Teilchen dieser Familie, das in der Natur in großen Mengen durch sogenannte Teilchenschauer in der Atmosphäre. Die dritte und deutlich schwerste Generation besteht aus dem Top-Quark, dem schwersten aller Quarks, mit einer Masse von $(172,76 \pm 0,30)$ GeV [11] und dem Bottom-Quark. Damit ist es fast 40000 mal schwerer als das Up-Quark. Das Bottom-Quark, das eine schwere Kopie des Down-Quark ist, ist im Vergleich nur rund 1000 mal so schwer, mit einer Masse von $(4,18^{+0,03}_{-0,02})$ GeV [11]. Das Tauon ist etwa 3000 mal schwerer als das Elektron mit einer Masse von $(1776,86 \pm 0,12)$ MeV [11]. Die Neutrinomasse wird theoretisch zwar als 0 angenommen, allerdings muss sie von null verschieden sein, da Neutrinos ineinander oszillieren können [12].

Die Eichbosonen sind die Austauschteilchen der drei Wechselwirkungen im Standardmodell. Sie haben einen ganzzahligen Spin und tragen bis auf das W^\pm -Boson keine Ladung. Das Photon ist das masselose Austauschteilchen der elektromagnetischen Wechselwirkung. Das Gluon, das ebenfalls masselos ist, vermittelt die starke Wechselwirkung zwischen den Quarks und Gluonen. Quarks und Gluonen tragen Farbladungen: rot, blau bzw. grün. Das Gluon überträgt die Farbladung und trägt jeweils eine Farbladung, die übertragen wird und eine Anti-Farbladung entsprechend der Farbe, die ersetzt wird [12]. Aufgrundessen gibt es 8 Gluonen.

Das neutral geladene Z^0 -Boson, sowie die W^\pm -Bosonen, welche eine positive oder negative Ladung tragen und somit auch elektrische Ladung vermitteln können, sind die schwersten Austauschteilchen. Die große Masse macht die schwache Kraft sehr kurzreichweitig [12].

3.1.2 Wechselwirkungen

Das Standardmodell beschreibt drei der vier grundlegenden Kräfte der Natur. Lediglich die Gravitationskraft kann nicht in Einklang mit dem Konzept gebracht werden. In dieser Theorie werden zudem die elektromagnetische und die schwache Wechselwirkung in einer einzelnen Wechselwirkung zusammengefasst: die elektroschwache Wechselwirkung. Dadurch vereint das Standardmodell die Theorien der Quantenelektrodynamik (QED) und die Fermi-Theorie [12].

Richard Feynman war es, der durch seine “Quantentheorie der Elektrodynamik” die Regeln für veranschaulichende Feynman-Diagramme einführte. Darin gibt es drei Arten von Linien, mit denen verschiedene Teilchentypen unterschieden werden (Abb. 8) [12]. Das klassische Austauschteilchen der elektromagnetischen Wechselwirkung ist das Photon. Es koppelt allerdings nur an Teilchen, die eine Ladung tragen, das heißt das Photon überträgt Energie und Impuls von einem Teilchen an seinen Wechselwirkungspartner, wenn dieser elektrische Ladung trägt.

Feynman-Diagramme beschreiben das Wechselwirken und Streuen einlaufender Teilchen und stellen Wechselwirkungen dar, die zu einer Menge von auslaufenden Teilchen führen. Es können zudem auch Loops entstehen, in denen virtuelle Teilchen erzeugt und durch ein entsprechendes Austauschteilchen vernichtet werden. Die Punkte werden durch den Propagator verbunden. Die fermionischen Quarks werden durch Linien dargestellt, deren Pfeile die Richtung der Zeit angeben. Photonen werden als gewellte Linie dargestellt während die restlichen Eich-Bosonen durch gestrichelte Linien dargestellt werden.

Glukonen werden durch gelockte Linien dargestellt.

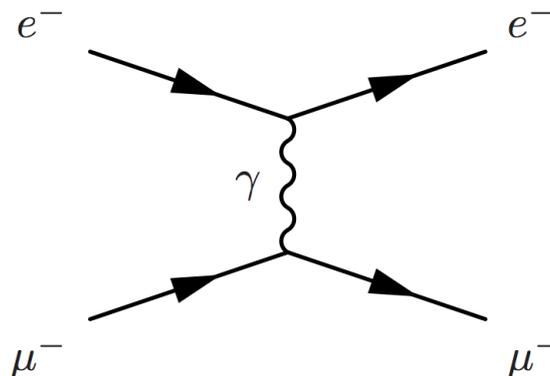


Abbildung 8: Feynman-Graph der durch ein virtuelles Photon vermittelten Wechselwirkung eines Elektrons mit einem Myon. Diese Darstellung zeigt die trivialste Form einer Streuung.

Ein Feynman-Graph stellt allerdings nur ein Szenario einer solchen Wechselwirkung dar, durch Loop-Diagramme gibt es weiter mögliche Abläufe des gleichen Zerfalls. Die Summe aller möglichen Diagramme wird benötigt, um die Wahrscheinlichkeit des dargestellten Prozesses zu berechnen. Häufig genügt es, die entsprechenden Pfadintegrale nur bis zu einer gewissen Ordnung von Loop-Diagrammen zu berechnen. Damit lässt sich das Verzweungsverhältnis eines Teilchenzerfalls berechnen, also die Wahrscheinlichkeit, dass ein bestimmtes Teilchen in diesem bestimmten Zerfallskanal zerfällt [12].

Die W^\pm -Bosonen und das Z -Boson sind die schwersten Austauscheteilchen im Standardmodell, die unter anderem Anlass für die Entwicklung des Higgs-Mechanismus waren. Alle Materieteilchen des Standardmodells nehmen an der schwachen Wechselwirkung teil. Die Besonderheit an der schwachen Wechselwirkung, ist ihre Eigenschaft Ladung zu transportieren. Daher vermitteln bei geladenen Strömen die W^\pm -Bosonen und bei neutralen Strömen die Z^0 -Boson.

Die schwache Kraft wurde erstmals 1933 durch Fermi formuliert, um den β -Zerfall zu erklären. Bereits 1930 postuliert Pauli die Existenz von Neutrinos als Wechselwirkungspartner im β -Zerfall (Abb. 9). Die endgültige Formulierung erfolgte durch Weinberg, Glashow und Salam in den 1960er Jahren bei der Formulierung des Standardmodells [12].

In den Dubletts der Leptonen können die Neutrinos der drei Generationen ineinander oszillieren. Dieser Effekt tritt auch bei den Quarks auf und für dessen Berücksichtigung in der Theorie wurde die CKM-Matrix (Cabibbo-Kobayashi-Maskawa) 1973 eingeführt [14]. Die CKM-Matrix beschreibt die Mischung der Flavourzustände (Gl. 2). Die Betragsquadrate der Einträge $|V_{ij}|^2$ stellen die Übergangswahrscheinlichkeit eines $-\frac{1}{3}$ -geladenen Quarks in ein $+\frac{2}{3}$ -geladenen Quark dar.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2)$$

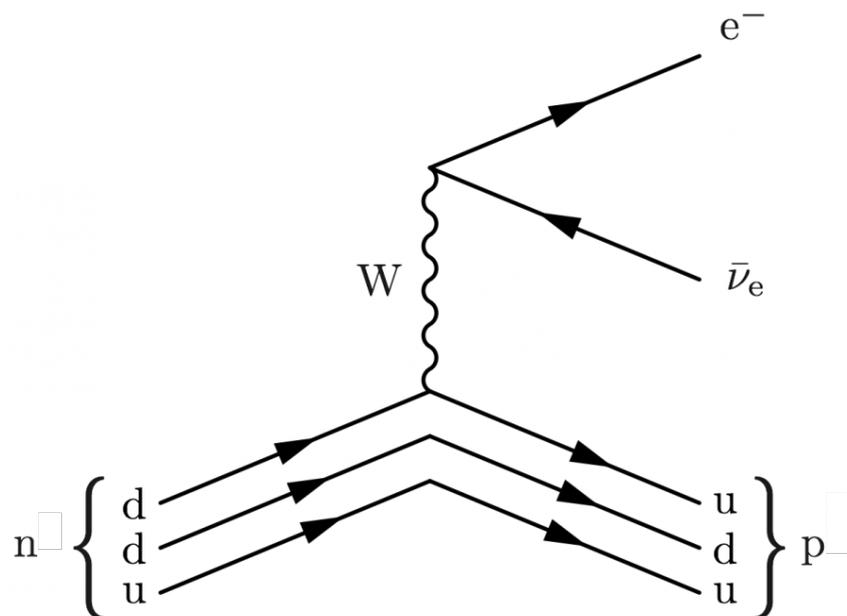


Abbildung 9: Darstellung des Beta-Zerfalls durch den Austausch eines negativ geladenen W-Bosons in ein Elektron und Elektron-Antineutrino.

Die starke Wechselwirkung wird durch acht verschiedene Gluonen als ihr Austauscheteilchen vermittelt. Quarks besitzen eine der drei Farbladungen blau, grün und rot. Um zwischen zwei Quarks vermitteln zu können, muss das Gluon eine entsprechende Kombination aus einer Farbladung und einer Anti-Farbladung besitzen. Die unterschiedlichen Quarkkombinationen müssen "weiß", entsprechend also farbneutral geladen sein. Daher bestehen Mesonen aus einem Quark-Antiquark-Paar mit entgegengesetzter Farbladung (Abb. 10) und Baryonen aus 3 Quarks, die paarweise verschiedene Farben tragen. Die zugrundeliegende Theorie wird Quantenchromodynamik (QCD) genannt [12]. Quarks können nicht als einzelnes Elementarteilchen existieren (Confinement). Der sogenannte Farbstring eines Quarkpaares erzeugt bei dem Versuch, das Paar durch entsprechend hohe Energie zu trennen, ein neues Quark-Antiquark-Paar. Gluonen sind aufgrund ihrer Eigenschaften auch in der Lage, mit sich selbst wechselzuwirken.

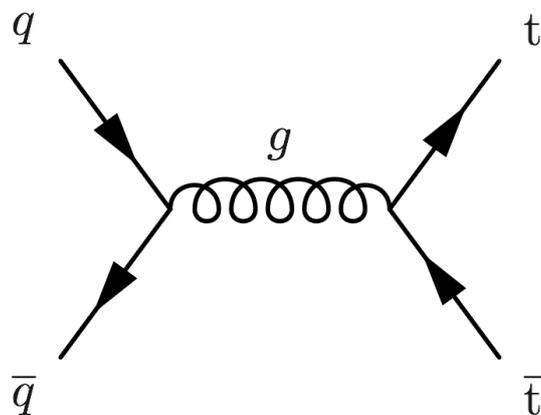


Abbildung 10: Darstellung der Anihilation zweier Quarks unter Verwendung eines Gluons zur Erzeugung eines Top-Antitop-Quarkpaares.

3.1.3 Der Higgs-Mechanismus

Der Higgs-Mechanismus wurde 1964 fast zeitgleich von drei unabhängigen Forschergruppen entwickelt und veröffentlicht [4] [5]. Das Konzept erklärte das Auftreten von massereichen Austauschteilchen in der schwachen Wechselwirkung. Darin wird ein Teilchen postuliert, das an die Masse der Teilchen koppelt und das Austauschteilchen der Masse darstellt. Dem liegt ein skalares Higgs-Feld zugrunde in welchen sich alle Teilchen bewegen. Je stärker sie an das Higgs-Boson koppeln, umso größer ist ihre entsprechende Masse.

2012 veröffentlichte die ATLAS- und CMS-Kollaborationen die Entdeckung eines bisher unbekanntes Teilchens, dessen Eigenschaften und Masse dem postulierten Higgs-Teilchen entsprechen [5] [4]. Seine Masse konnte durch vorhergehende Experimente eingegrenzt werden. Für die Entdeckung erhielten Peter Higgs und François Englert den Nobelpreis.

3.2 Die Fisher-Diskriminante

Die lineare Diskriminanzfunktion wurde 1936 von R. A. Fisher entwickelt [6] und dient der Trennung zweier Klassen n -dimensionaler Parameter, indem die Parameter auf eine entsprechend des Fisherschen Kriterium gewählte $n - 1$ dimensionale Mannigfaltigkeit projiziert werden, um die Trennbarkeit der Klassen zu maximieren. Sie ist die einfachste Form einer Diskriminanzfunktion, da sie linear von allen Parametern abhängt [6].

3.2.1 Ausgangsproblem

Um die Fisher-Diskriminante berechnen zu können, muss die untersuchte Menge von Parametern gewissen Voraussetzungen genügen: Die wichtigste Voraussetzung ist, dass die Daten in zwei Klassen eingeteilt werden können. Im Kontext dieser Arbeit sollen verschiedene physikalische Messgrößen eines typischen Kollisionsexperiments in die Klassen Signal und Untergrund aufgeteilt werden.

Für die Diskriminanzfunktion ist die häufigste Basis die Trennung von Untergrund und Signal in einer Verteilung durch einen entsprechenden Schnitt auf mehrere Variablen, welcher als Entscheidungsgrenzwert dient. Dabei kann nie eine perfekte Trennung von Signal und Untergrund erzielt werden. Ziel ist die optimale Unterdrückung des Untergrundes im Verhältnis zu den Signalereignissen. Entsprechend der Anzahl der Parameter, die gleichzeitig optimiert werden sollen, ist diese Trennebene eine Hyperebene geringerer Dimension als der Raum der durch die Parameter aufgespannt wird (Abb. 11). Ziel dieser Projektion ist die Reduzierung der Daten um die Trennbarkeit beider Klassen zu maximieren und die unerwünschte Klasse zu unterdrücken [6].

Wie im Verlauf dieser Arbeit zu sehen ist, werden unterschiedliche Messgrößen, die Parameter für das Fisher Kriterium, eingelesen. Dabei sollen Ereignisse separiert werden, die zu einer Vielzahl von anderen Zerfällen gehören als der zu untersuchende Zerfall. Zum Ermitteln der Koeffizienten für das Kriterium werden zunächst simulierte Daten verwendet, bei denen genau bekannt ist, welche Ereignisse Signale sind und welche zum Untergrund gehören. Die Form der Verteilungen von simulierten Signalen und Untergrund in Summe soll dabei mit den Messdaten gut übereinstimmen. Nach der Berechnung der Koeffizienten kann dann das Fischerkriterium auf die experimentellen Messdaten angewendet werden,

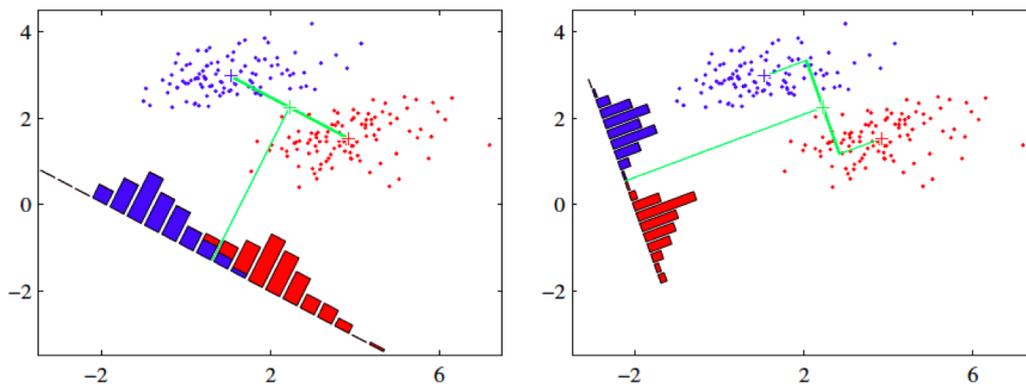


Abbildung 11: Einfaches Beispiel zur Veranschaulichung des Funktionsprinzips der Fisher-Diskriminante durch lineare Projektion [15].

um eine optimale Unterdrückung von Untergrund bei zeitgleicher Signaloptimierung zu erreichen.

3.2.2 Mathematisches Vorgehen

Das Ausgangsproblem für die Fisher-Diskriminante wird durch eine scheinbar triviale Gleichung dargestellt. Dabei können zwei Klassen unterschieden werden, die für die Berechnung der Diskriminanzstatistik von entscheidender Bedeutung sind. Dazu soll aus einem Vektor von Messdaten oder Parametern $\vec{x}^T = (x_1, \dots, x_{N_P})$ in einfachster Weise eine eindimensionale Statistik gebildet werden. Das Verhältnis der beiden Wahrscheinlichkeitsverteilungen $h(x|H_S)$ mit H_S als Signalthypothese und $h(x|H_U)$ mit H_U als Untergrundhypothese bildet die Teststatistik $t(x)$.

$$t(x) = \frac{h(x|H_S)}{h(x|H_U)} \quad (3)$$

Kern der Methode ist die Bestimmung der N_P Koeffizienten a_i die mit einer Linearkombination der Parameter die Statistik $t(\vec{x})$ bildet.

$$t(\vec{x}) = \sum_{i=1}^{N_P} a_i x_i = \vec{a}^T \vec{x} \quad (4)$$

dabei ist $\vec{a}^T = (a_1, a_2, \dots, a_{N_P})$ der transponierte Vektor der Koeffizienten jedes Parameters und i der jeweilige Parameter. Die Zahl unabhängiger Variablen gibt die Dimension der Statistik an. Die Koeffizienten werden mittels des arithmetischen Mittels M (Gruppenmittelpunkt) der beiden Klassen bestimmt. Dazu werden zunächst alle Parameter einer Klasse unter sich gemittelt, sodass man für den i -ten Parameter den arithmetischen Mittelwert $(M_k)_i$ und das arithmetische Mittel M aller Parameter und Klassen erhält.

$$(M_k)_i = \frac{1}{N_k} \sum_{j=1}^{N_k} x_{ij}$$

$$M = \frac{1}{N_p} \sum_{i=1}^{N_p} (M_S + M_U)_i$$
(5)

Der Index k beschreibt dabei die jeweilige Klasse: Signal (S), Untergrund (U). Dabei steht N_k für die Zahl der Ereignisse innerhalb einer Klasse und N_p für die Zahl von Parametern. Zudem werden die Kovarianzmatrizen S_W für die Kovarianz innerhalb (englisch: within; daher der Index W) einer Klasse und S_B für die Kovarianz zwischen (englisch: between; daher der Index B) den zwei Klassen ermittelt. Diese Matrizen beschreiben ein Maß von Streuung und Korrelation des Parameters einer Klasse in Bezug zu seinem arithmetischen Mittel. Die Kovarianzmatrix S_B zwischen den Klassen wird durch ihren Gruppenmittelpunkt M sowie M_k bestimmt.

$$(S_W)_k = \sum_{i=1}^{N_p} \sum_{j=1}^{N_k} N_k (x_{ij} - (M_k)_i) (x_{ij} - (M_k)_i)^T$$

$$S_B = \sum_{i=1}^{N_p} N_k ((M_k)_i - M) ((M_k)_i - M)^T$$
(6)

Dabei steht N_k für die Zahl der Messpunkte innerhalb einer Klasse eines Parameters, N_p für die Zahl der Parameter und i für den jeweiligen Parameter. Für ein System aus zwei Klassen erhält man zwei Kovarianzmatrizen der jeweiligen Klasse. Die elementweise Summe aus beiden Kovarianzmatrizen bildet die Kovarianzmatrix S_W . Mittels der Inversen der Kovarianzmatrix S_W lässt sich nun aus den Mittelwerten innerhalb der Klassen M_S und M_U der Koeffizientenvektor \vec{a} berechnen.

$$\vec{a} = k \cdot S_W^{-1} (M_S - M_U) + f$$
(7)

Die damit errechneten Koeffizienten a_i können nun in die zu Beginn eingeführte Gleichung eingesetzt werden und ergeben damit die Teststatistik $t(\vec{x})$. Zu beachten ist aber, dass diese Koeffizienten noch von einem relativen gemeinsamen Faktor k und einer additiven Verschiebung f abhängen können [6].

3.3 Boosted Decision Trees

Ein Boosted Decision Tree (BDT) [7] ist eine Methode des Maschinellen Lernens, in der die Ereignisse eines Datensatzes in zwei Klassen kategorisiert werden. Dabei muss eine von vielen vorher aus dem Trainingsdatensatz eingelesenen Variablen eines Ereignisses einen größeren bzw. kleineren Betrag als ein Schnittwert besitzen. Der einfache Schnittwert wird optimal auf Basis des Datensatzes ausgewählt. Dies geschieht mit allen Ereignissen, die währenddessen gleich gewichtet werden. Dadurch wird der Datensatz in zwei Klassen eingeteilt, bspw. signalähnlich und untergrundähnlich [7].

Im nächsten Schritt werden diese beiden Teildatensätze jeweils einem neuen Kriterium für eine der anderen Variablen auf gleiche Weise wie schon im ersten Knoten unterzo-

gen. Es entstehen vier Teildatensätze (Baskets) und zwei Knoten. Dieser Prozess des Kategorisierens und Gewichtens wird fortgesetzt bis bspw. das Verhältnis der Zahl der Ereignisse der Teildatensätze zur gesamten Zahl an Ereignissen einen Grenzwert unterschreitet. Durch das fortschreitende Aufspalten in n Schritten entstehen 2^n Teildatensätze (Baskets) und 2^{n-1} Knoten.

Die Option Adaptive Boost (AdaBoost) [7] ermöglicht es, dass im ersten Knoten des BDT, also dem Punkt in dem der Datensatz zum ersten Mal kategorisiert wird, die Gewichte der Ereignisse erhöht werden, die einer falschen Klasse zugeordnet wurden [7]. Daher ist es wichtig, dass der Datensatz (Trainingsdatensätze für Signal und Untergrund), mit dem der Decision Tree trainiert wird, in Signal und Untergrund unterschieden werden kann. Die durch AdaBoost gewichteten Ereignisse bilden einen neuen Trainingsdatensatz, wobei nun die falsch kategorisierten Ereignisse stärker gewichtet werden. Mit diesem geänderten Trainingsdatensatz wird ein neuer BDT gebildet und jeder neue BDT basiert auf den Gewichten derer davor.

Solche Decision Trees werden zu hunderten nacheinander trainiert, um am Ende ein Modell zu erhalten, einen BDT, der die Ereignisse möglichst präzise in die zwei Klassen aufteilen kann. Nach jedem Durchlauf werden die durch AdaBoost ermittelten Gewichte auf den Trainingsdatensatz angewandt bevor der nächste BDT erstellt wird. Nach Beendigung des Durchlaufs sind hunderte BDTs entstanden, die final auf verschiedene Weisen miteinander kombiniert werden können. Daher ist es möglich, eine nichtlineare Hyperfläche zu finden, die den Untergrund vom Signal trennt.

3.4 Grenzwertoptimierungskurve

Bei einer Grenzwertoptimierungskurve, kurz ROC für Receiver Operator Characteristic, wird die Unterdrückungsrate $1 - \varepsilon_{bkg}$ gegen die Signaleffizienz ε_{sig} aufgetragen. Die daraus resultierende Funktion gibt die Unterdrückung des Untergrunds für eine bestimmte Signaleffizienz an. Bei verbesserter Unterdrückung des Untergrunds und Separation des Signals formt sich die Kurve mit einer stärkeren Krümmung in dem Bereich nahe dem Verhältnis von kompletter Unterdrückung ($1 - \varepsilon_{bkg} = 1$) und voller Signaleffizienz ($\varepsilon_{sig} = 1$). Ein Maß für die Güte eines ROC-Graphen ist die Fläche unter dem Graphen, die aufgrund einer diskreten Verteilung nur approximiert werden kann. Eine ideale Grenzwertoptimierungskurve existiert allerdings nicht.

Die Signaleffizienz ε_{sig} errechnet sich aus dem Verhältnis von der Anzahl an Signalereignissen $N_{sig, cut}$, die einen Schnitt auf eine Variable erfüllen, zur totalen Anzahl an Signalereignissen N_{sig} . Genauso errechnet sich die Untergrundeffizienz ε_{bkg} aus dem Verhältnis der selektierten Untergrundeignisse $N_{bkg, cut}$ zu den totalen Untergrundeignissen N_{bkg} . Mittels der Untergrundeffizienz ε_{bkg} kann die Unterdrückungsrate $1 - \varepsilon_{bkg}$ berechnet werden. In einem ROC-Graphen ist die Unterdrückungsrate $1 - \varepsilon_{bkg}$ der Signaleffizienzen ε_{sig} als Funktion des Schnittes aufgetragen. Anhand des Graphen können konkrete Arbeitspunkte mit gewünschter Signaleffizienz ε_{sig} abgelesen werden. Trägt man die Signaleffizienz gegenüber dem Kriterium auf, lassen sich damit einer Effizienz ε_{sig} entsprechende Schnitte auf die zugehörige Verteilung ermitteln.

4 Der untersuchte Zerfall

Wenn in Teilchenbeschleunigerexperimenten Quarks produziert werden, beginnen diese aufgrund des Confinements zu hadronisieren. Es bilden sich nach Kollisionen Mesonen und Baryonen, die nach kurzer Lebensdauer wieder zerfallen. Im Detektor ist dann nur noch ein Schauer von Teilchen feststellbar und nicht mehr das ursprüngliche Quark oder Gluon.

Der in dieser Arbeit untersuchte Zerfall, ist der eines B^+ -Mesons in ein J/ψ -Meson, einer Quark-Anti-Quark Resonanz des Charm-Quarks, und ein positiv geladenes K^+ -Meson, bestehend aus einem Anti-Strange-Quark und einem Up-Quark (Abb. 12). Das B^+ -Meson ist ein Quark-Antiquark-Zustand aus einem Anti-Bottom-Quark und einem Up-Quark. Des Weiteren werden Zerfälle von Bottom-Anti-Bottom-Quarkresonanzen betrachtet. Für die weitere Analyse in dieser Arbeit soll aber im wesentlichen der Zerfall des B^+ -Mesons betrachtet werden. Wenn zwei Protonen kollidieren entstehen Quarks und Gluonen und die Quarks strahlen wiederum Gluonen ab. Aufgrund des Confinements hadronisieren die b -Quarks, unter anderem zu einem B^+ -Meson, welches dann nach kurzer Lebensdauer im zweiten Vertex zerfällt. Es zerfällt mit einer mittleren Lebensdauer von $\tau_B = 1,638 \cdot 10^{-12} s$ [11] in weitere Mesonen und Leptonen.

Ein möglicher Zerfall resultiert in einem J/ψ -Meson und einem K^+ -Meson. Das Verhältnis der Verzweigungsverhältnisse $\mathcal{B}(B_u^+ \rightarrow J/\psi\pi^+)/\mathcal{B}(B_u^+ \rightarrow J/\psi K^+) = (5,0 \pm 1,8)\%$ für den $J/\psi K^+$ -Zerfall ist vergleichsweise klein [16]. Das J/ψ -Meson zerfällt wiederum in zwei Myonen, die dann im Myonen-Spektrometer detektiert werden.

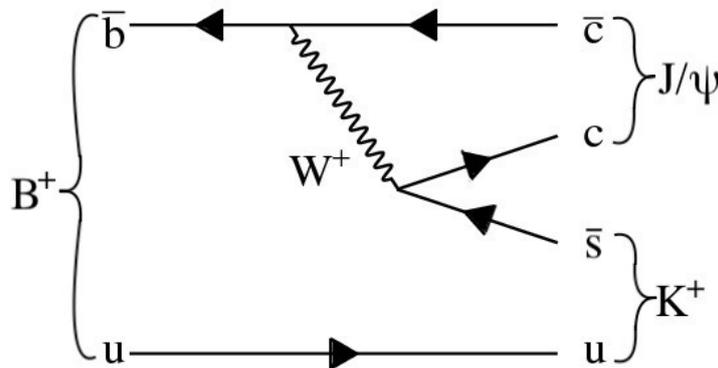


Abbildung 12: Der Zerfall eines B^+ -Mesons in ein J/ψ -Meson und ein K^+ -Meson. Das J/ψ -Meson zerfällt weiter in zwei Myonen.

Das J/ψ -Meson wurde 1974 fast zeitgleich von zwei Forschergruppen unabhängig voneinander in den Vereinigten Staaten entdeckt. Es bestätigte mit seiner Entdeckung die Existenz des 4. Quarks, dem Down-Quark ähnlichen Partner des Strange-Quarks, dem Charm-Quark, da es aus einem Charm-Quark-Anti-Charm-Quark-Paar besteht. Es zeichnet sich besonders durch seine ungewöhnlich schmale Zerfallsbreite von $(92,9 \pm 2,8) \text{ keV}$ [11] aus, wegen welcher es eine um rund 3 Größenordnungen längere Lebensdauer als andere Quark-Resonanzen besitzt.

4.1 Die untersuchten Größen

Während der pp -Kollisionen im ATLAS-Detektor werden die deponierten Energien, die Spurpunkte und die dazugehörigen Zeiten der einzelnen Ereignissen erfasst. Ausgehend davon werden dann im Anschluss weitere Größen rekonstruiert, von denen sich einige besonders gut zur Trennung der gesuchten Zerfallsprozesse eignen und andere nicht. Um die folgenden Größen zu verstehen, müssen zunächst einige grundlegende Größen eingeführt werden.

Der Transversalimpuls p_T ist ein wichtiges Maß, da dieser in jeder pp -Kollision am LHC einen Anfangswert von 0 hat und erhalten bleibt. Die Ursache findet sich in der experimentellen Anordnung aufgrund des Boosts des Parton-Parton-Systems in Strahlrichtung. Daher werden für diese Analyse die Komponenten planar zu x - y -Ebene betrachtet. Der Transversalimpuls berechnet sich aus der x -Komponente p_x und der y -Komponente p_y des Impulses, also der Projektion des Impulsvektors auf die entsprechende Achse des euklidischen Raumes. Der Transversalimpuls ist definiert als

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (8)$$

Die Richtung einer Teilchentrajektorie kann durch zwei Winkel charakterisiert werden. Dazu wird der Polarwinkel Θ zwischen Strahlachse und Trajektorie angegeben. Der Azimutalwinkel ϕ beschreibt den Winkel um die Strahlachse bezüglich der x -Achse des Koordinatensystems. Mittels des Polarwinkels kann nun die Pseudorapidität berechnet werden. Die Pseudorapidität η ist ein anderes Maß für den Winkel von Trajektorien, das den Vorteil hat, dass Differenzen dieser Größe Lorentz-invariant sind:

$$\eta = -\ln \left(\tan \left(\frac{\Theta}{2} \right) \right). \quad (9)$$

Eine letzte wichtige Größe ist der Abstand ΔR zwischen zwei Spuren mit den Pseudorapiditäten η_i und Azimutalwinkeln ϕ_i :

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (10)$$

Hier ist $\Delta\eta$ der Unterschied von Pseudorapiditäten der zwei Spuren und $\Delta\phi$ die Differenz der Winkel [17]. Im Zentrum dieser Untersuchung sollen besonders drei Messgrößen stehen, die im Folgenden hierarchisch ihrer Bedeutung zur Separation von Signal und Untergrund nach gegliedert sind:

1. Transversale Zerfallslänge L_{xy} : Der Betrag der Projektion von $\Delta\vec{x}$, dem Abstand des primären Zerfallsvertex zum sekundären Vertex, auf die Achse, die durch den Transversalimpuls des B-Mesons \vec{p}_T^B definiert wird [17].
2. Projektionswinkel α_{2D} : Der Projektionswinkel zwischen $\Delta\vec{x}$ und dem Transversalimpuls des B-Mesons \vec{p}_T^B auf der zur Strahlachse orthogonalen x - y -Ebene (Abb. 13).
3. Isolation $I_{0.7}$: Die Isolation wird durch den Quotienten vom Betrag des Transversalimpuls des B-Mesons $|\vec{p}_T^B|$ zu der Summe aus $|\vec{p}_T^B|$ und der Summe der Absolutbeträge aller weiteren Transversalimpulse $|\vec{p}_T^i|$ berechnet, $I_{0.7} = \frac{|\vec{p}_T^B|}{|\vec{p}_T^B| + \sum |\vec{p}_T^i|}$, die nicht aus dem B-Vertex stammen, innerhalb eines durch ΔR , dem Konusradius, definierten Öffnungskegel um den Impuls \vec{p}_T^B am primären Vertex (Abb. 14). Die Beträge der Transversalimpulse der Spuren müssen einen minimalen Wert überschreiten

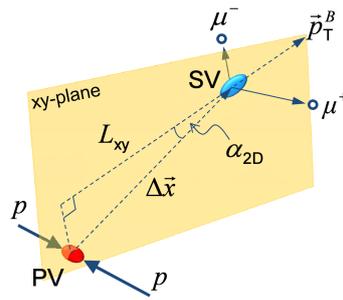


Abbildung 13: Die Projektion L_{xy} der Zerfallslänge $\Delta\vec{x}$ auf den auf die Transversalebene projizierten Impuls \vec{p}_T^B mit dem eingezeichneten Projektionswinkel α_{2D} [17].

($|\vec{p}_T^B| > 0,5 \text{ GeV}$). Der Öffnungswinkel des Kegels wird durch ΔR definiert. Hier ($I_{0,7}$) beträgt $\Delta R = 0,7$ [17].

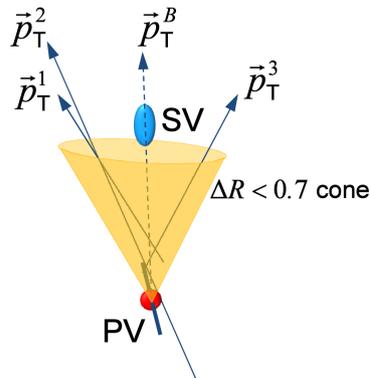


Abbildung 14: Der Öffnungskegel am primären Vertex um $|\vec{p}_T^B|$, der der Auswahl der Spuren dient, die zur Berechnung der Isolation des B-Mesons verwendet werden [17].

5 Optimierung

Im folgenden Kapitel sollen unterschiedliche Methoden verwendet werden, um den Untergrund zu unterdrücken, ohne dabei zu viele Signalereignisse zu verlieren. Dazu werden drei Methoden verglichen:

- i) Eindimensionaler Schnitt: Schnitt auf einen einzelnen Parameter (Kap. 5.2).
- ii) Fisher-Diskriminanzkriterium: optimale Linearkombination mehrerer Parameter zur Verbesserung der Separabilität (Kap. 5.3).
- iii) Boosted Decision Trees: auf TMVA (Toolkit for Multivariate Data Analysis) basierendes multivariates Verfahren (Kap. 5.4).

Zunächst werden die Methoden auf den in Kapitel 4 eingeführten drei Größen ($|\alpha_{2D}|, I_{0.7}, L_{xy}$) trainiert und validiert, ehe sie darauffolgend auf Daten des ATLAS-Experiments angewendet werden. Zur Vergleichbarkeit der Methoden werden während der Validation ROC-Graphen erstellt, um die Effektivität der Methoden zu vergleichen.

5.1 Verwendete Datensätze

Für das Trainieren der Methoden ist es wichtig, dass die Monte-Carlo (MC) Daten klar in Signal und Untergrund unterschieden werden können. Die Signalereignisse sind simulierte $B^+ \rightarrow J/\psi K^+$ Zerfälle mit $J/\psi \rightarrow \mu^+ \mu^-$. Aus den Spuren der Myonen können die Zerfälle der J/ψ -Mesonen und, zusammen mit den Spuren der K^+ -Mesonen dann die Zerfälle der B^+ -Mesonen rekonstruiert werden. Um reine Signalereignisse in diesen Simulationsdaten einzulesen, wird auf den Parameter `isSignal > 0` geschnitten. Der Parameter `isSignal` enthält die Information, ob die rekonstruierten Spuren mit dem Impuls der Partonen überlappen. Dadurch enthalten diese NTupel nur Ereignisse mit B^+ -Kandidaten. B^+ -Kandidaten sind Ereignisse, bei denen die invariante Masse von $\mu^+ \mu^- K^+$ der erwarteten Massenverteilung des B^+ -Mesons entspricht.

Die Untergrundereignisse stammen aus dem MC-Datensatz $b\bar{b} \rightarrow \mu^+ \mu^- X$, bei dem wieder die invariante Masse von zwei Myonen und einem Kaon $m_{\mu\mu K}$ betrachtet wird. Er enthält Zerfälle der Form $b\bar{b} \rightarrow \mu^+ \mu^- X$. Hier wurde ein Schnitt auf den `hasTruthB`-Marker gesetzt, welcher $B^\pm \rightarrow J/\psi K^\pm$ Zerfälle ausschließt, um reine Untergrundzerfälle einzulesen.

Zusammen bilden diese NTupel den MC-Datensatz zur Implementierung und unabhängigen Überprüfung der Methoden. Ein NTupel ist eine Speicherform, die vergleichbar mit einer Liste von Arrays ist. Jedem Ereignis wird ein Array von Parameterwerten zugeordnet. Für das Trainieren der Methoden werden die Ereignisse aus einem NTupel eingelesen. Zum Testen und Trainieren wird der Datensatz in zwei unabhängige Teildatensätze aufgeteilt. Insbesondere für die Fisher-Diskriminante, sowie TMVA ist es wichtig, dass die Datensätze mit denen die Methoden trainiert werden, statistisch unabhängig von denen sind, mit denen sie getestet werden.

Neben diesen MC simulierten Datensätzen sollen die Methoden nach Implementierung und Validation auf den MC-Daten auf experimentelle Daten von ATLAS angewendet werden. Diese Daten stammen aus den Jahren 2015 und 2016 und enthalten Kandidaten für

den $B^\pm \rightarrow J/\psi K^\pm$ Zerfall (Tab. 1). Dafür wurde dieser Prozess mit dem $J/\psi \rightarrow \mu^+ \mu^-$ Zerfall aus den Spuren der Myonen in den Daten rekonstruiert. Diese Daten enthalten keine Signal oder Untergrund Markierung wie `isSignal` oder `hasTruthB`. Die Methoden werden hier angewendet, um den Untergrund optimal zu unterdrücken und Signalereignisse sichtbar zu machen.

Zerfallskanal bzw. Endzustand	Verwendung	Zahl der Ereignisse N_{total}	Zahl der Ereignisse N_k (Signal bzw. Untergrund)
$B^+ \rightarrow J/\psi K^+$ mit $J/\psi \rightarrow \mu^+ \mu^-$	Signal Datensatz	$1,108 \cdot 10^6$	$7,281 \cdot 10^5$
$b\bar{b} \rightarrow \mu^+ \mu^- X$ mit $J/\psi \rightarrow \mu^+ \mu^-$	Untergrund Datensatz	$1,612 \cdot 10^6$	$1,501 \cdot 10^6$
$\mu^+ \mu^- K^+$ Endzustand	ATLAS Daten	$1,792 \cdot 10^7$	-

Tabelle 1: Anzahl an Ereignissen in den unterschiedlichen Datensätzen, die für das Training, die Validation der beiden multivariaten Methoden und die Anwendung der Methoden verwendet wurden.

Die in Kapitel 4 eingeführten Größen, sowie die Masse des B^\pm -Mesonkandidaten, entsprechend die invariante Masse von $\mu^+ \mu^- K^\pm$, $m_{\mu\mu K}$, werden aus den Datensätzen eingelesen. Bei den MC-Datensätzen werden zudem noch die Werte der Signal- bzw. Untergrund-Marker `isSignal` bzw. `hasTruthB` eingelesen. Die in Abb. 15 und Abb. 16 dargestellten Histogramme zeigen die Monte-Carlo simulierten Datensätze für das Trainieren und Validieren der drei verschiedenen Methoden.

In Abb. 15 oben ist die Massenverteilung für die B^\pm -Kandidaten $m_{\mu\mu K}$ zu sehen. Zu beachten ist hier, dass die einzelnen Histogramme nicht normiert sind. Die ATLAS-Daten zeigen eine Ausprägung bei der erwarteten Masse des B-Mesons von $(5279,34 \pm 0,12)$ MeV, die auch in den Signaldatensätzen der MC-Daten zu sehen ist. Zur Verbesserung des Verhältnisses von Signal zu Untergrund sollen die Methoden Anwendung finden. In der Verteilung der Isolation (Abb. 15 unten) wird deutlich, dass bei sinkender Isolation der Anteil des Untergrund rapide steigt, sodass für eine gute Separation ein gewisser Anteil an Untergrundeffekten in Kauf genommen werden muss, um nicht zuviel Signal zu verlieren. Für den Projektionswinkel α_{2D} (Abb. 16 oben) ist die Verteilung symmetrisch. Für den Absolutbetrag $|\alpha_{2D}|$ sind nur die Ereignisse kleiner als ein Schnitt größer als null zu wählen. Dabei ist das Ziel, Arbeitspunkte auszuwählen, durch die der Untergrund möglichst stark unterdrückt wird, ohne einen zu großen Verlust bei der Signaleffizienz zu erleiden.

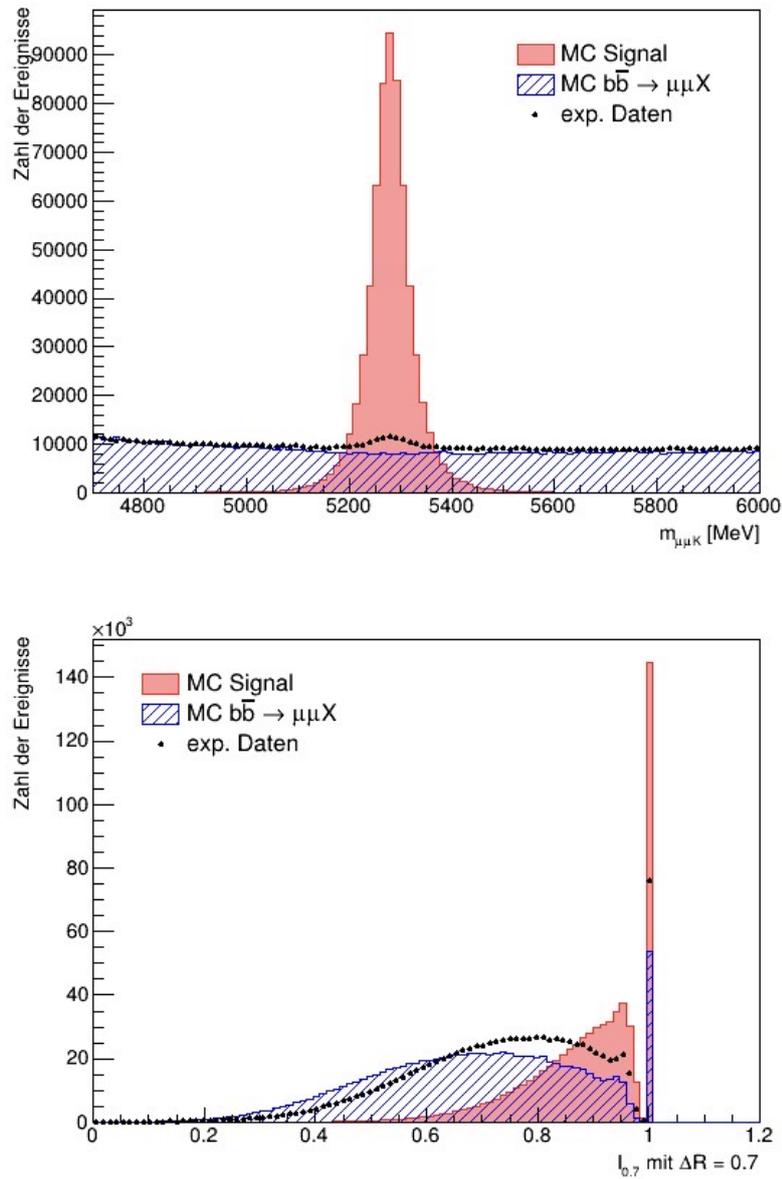


Abbildung 15: Die Monte-Carlo-Daten für Signal und Untergrund, sowie ATLAS-Daten von 2015 und 2016, der zwei Messgrößen: invariante Masse der Myonen und des Kaons $m_{\mu\mu K}$ (oben) und die Isolation $I_{0.7}$ (unten).

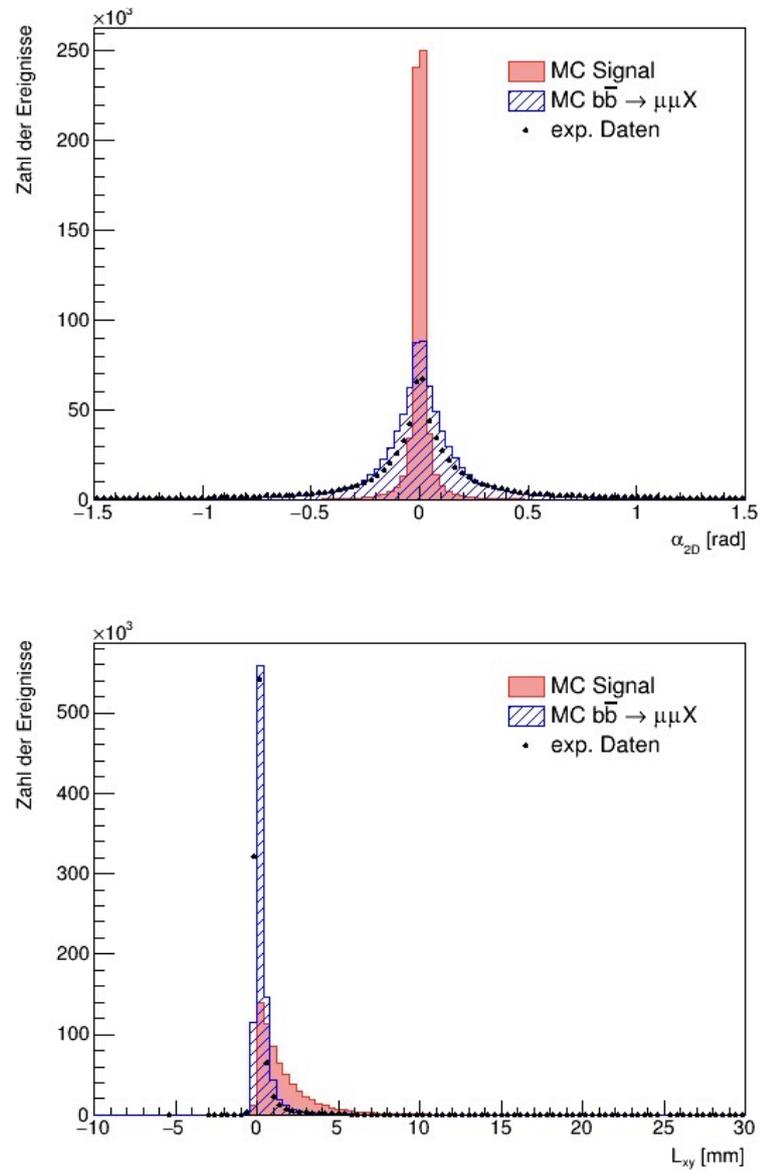


Abbildung 16: Die Monte-Carlo-Daten für Signal und Untergrund, sowie ATLAS-Daten von 2015 und 2016, der zwei Messgrößen: der Projektionswinkel α_{2D} (oben) und die transversale Zerfallslänge L_{xy} (unten).

5.2 Schnitt auf einen Parameter

Bei dieser Analyse geht es um den Vergleich der Trennungsstärke von Signal und Untergrund in Abhängigkeit von Schnitten, die auf einzelne Variablen der Messdaten gesetzt werden. Durch die Beschneidung des Betrachtungsintervalls und den Vergleich von Effizienzen wird analysiert, wie gut Teile des Untergrunds unterdrückt werden können ohne zu viele Signalereignisse zu verlieren. Mit dieser ersten Methode sollen nur Schnitte auf jeweils einen Parameter gesetzt werden.

5.2.1 Training auf Monte-Carlo-Daten

Bei dieser Methode wurden zunächst die Trainings-NTupel eingelesen, welche aus jeweils der Hälfte der Einträge des jeweiligen Signaldatensatzes und Untergrunddatensatzes bestehen. Für jedes Ereignis wird ein Masseneintrag nur dann in das Histogramm eingefüllt, wenn der zu diesem Ereignis zugehörige Parameter das gewählte Schnittkriterium erfüllt.

Am Beispiel des Projektionswinkels α_{2D} wird der Absolutbetrag des Winkels verwendet. Dieser muss kleiner als ein Grenzwinkel $|\alpha_{2D}^{cut}|$ sein, der in Intervallschritten von 0,05 rad im Intervall von 0 bis π vergrößert wird. Während jedes Schritts wird die Anzahl der Signalereignisse bzw. Untergrundereignisse innerhalb des Intervalls, also kleiner als $|\alpha_{2D}^{cut}|$, gezählt und das jeweilige Verhältnis zur absoluten Anzahl an eingelesenen Signal- bzw. Untergrundereignissen berechnet. Dadurch erhält man Werte für die Signal- und Untergrundeffizienz; mittels letzterer lässt sich die Unterdrückungsrate $1 - \varepsilon_{bkg}$ berechnen und in einem ROC-Graphen gegen die Signaleffizienz auftragen (Abb. 17). Dadurch erhält man die typische Form einer Grenzwertoptimierungskurve (ROC).

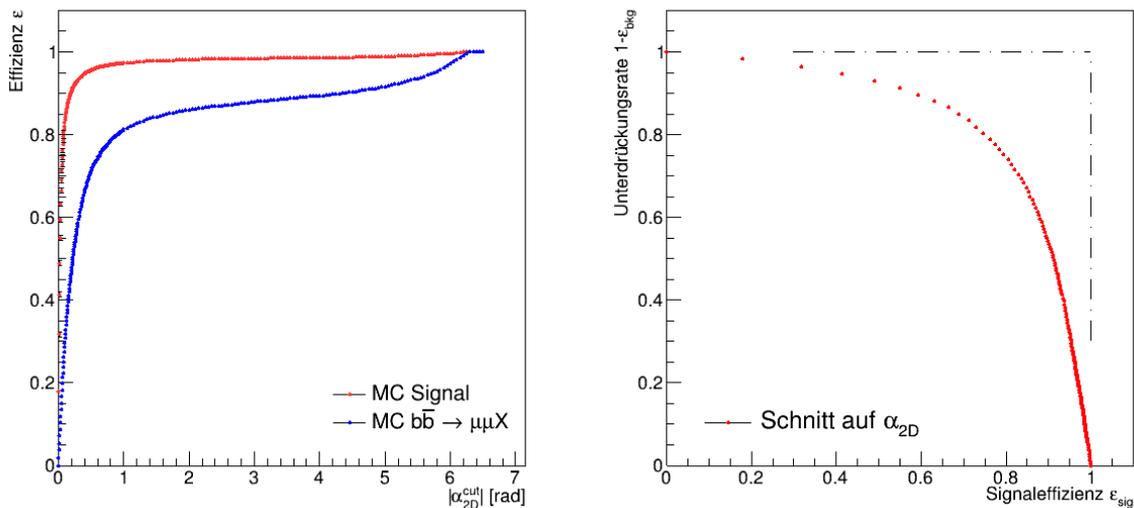


Abbildung 17: Links: Errechnete Effizienzen für Signal ε_{sig} (rot) und Untergrund ε_{bkg} (blau). Rechts: Grenzwertoptimierungskurve (ROC) für steigenden Schnitt auf $|\alpha_{2D}^{cut}|$.

Für die Isolation $I_{0.7}$ läuft der Schnittwert $I_{0.7}^{cut}$ in Schritten von 0.05 von 1 sukzessive bis 0. Es werden nur Ereignisse akzeptiert deren Isolation $I_{0.7}$ größer als $I_{0.7}^{cut}$ sind. Bei den

zugehörigen Effizienzgraphen fällt auf, dass die Signal- und Untergrundeffizienz nicht bei 0 starten (Abb. 18). Eine Isolation von eins ergibt sich immer dann, wenn nur Spuren des B^\pm Zerfalls im Öffnungskegel liegen. Daher gibt es auch für eine Isolation von 1 Ereignisse, die zu einer Signal- bzw. Untergrundeffizienz beitragen.

Bei der transversalen Zerfallslänge L_{xy} steigt der Schnittwert von 0 bis 20 in Intervallen der Breite 0,2 mm. Es werden nur Ereignisse akzeptiert, deren transversale Zerfallslänge L_{xy} kleiner als L_{xy}^{cut} ist. Auch hier erhalten wir einen ROC-Graphen (Abb. 19).

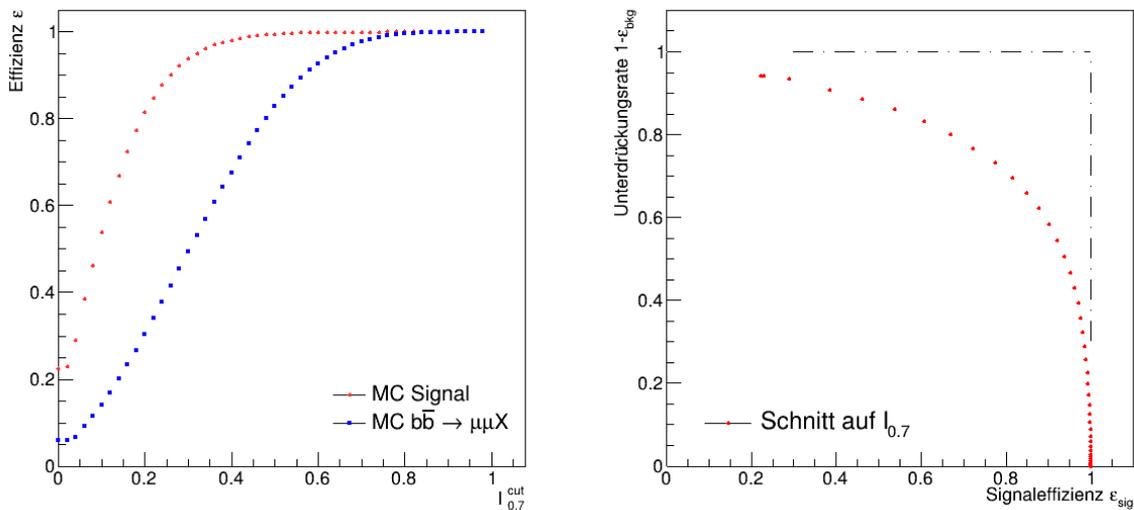


Abbildung 18: Links: Errechnete Effizienzen für Signal ϵ_{sig} (rot) und Untergrund ϵ_{bkg} (blau). Rechts: Grenzwertoptimierungskurve (ROC) für steigenden Schnitt auf $I_{0.7}^{cut}$.

In Abb. 20 wird besonders im direkten Vergleich deutlich, wie stark sich die Form des ROC-Graphs je nach Wahl des Parameters, für den ein Schnittkriterium eingeführt wurde, verändert. Der beste ROC-Graph ergibt sich im Vergleich der drei Parameter für den Projektionswinkel $|\alpha_{2D}|$. Der Parameter $|\alpha_{2D}|$ ist am besten geeignet, da die AUC (Area under Curve, engl. Fläche unter dem Graphen) für diese Kurve am größten ist. Sie ist ein Maß für die Güte einer Grenzwertoptimierungskurve und liegt für eine ideale Kurve, die bei jeder Signaleffizienz ϵ_{sig} eine Untergrundunterdrückungsrate von 1 besitzt, bei 1. Daher soll der zugehörige ROC-Graph von $|\alpha_{2D}|$ mit den anderen Methoden verglichen werden.

5.2.2 Anwendung auf reale Messdaten

Nach der Analyse der Parameter hinsichtlich ihrer Unterdrückungsrate und Signaleffizienz bei unterschiedlichen Schnittwerten, werden nun zwei Arbeitspunkte gewählt. Bei einer Signaleffizienz von 80 Prozent bzw. 90 Prozent, werden die zugehörigen Schnittwerte für die Parameter in den Effizienzgraphen abgelesen. Mittels dieser Werte können nun die realen Messdaten ausgewertet werden. Dafür werden entsprechende NTupel, die

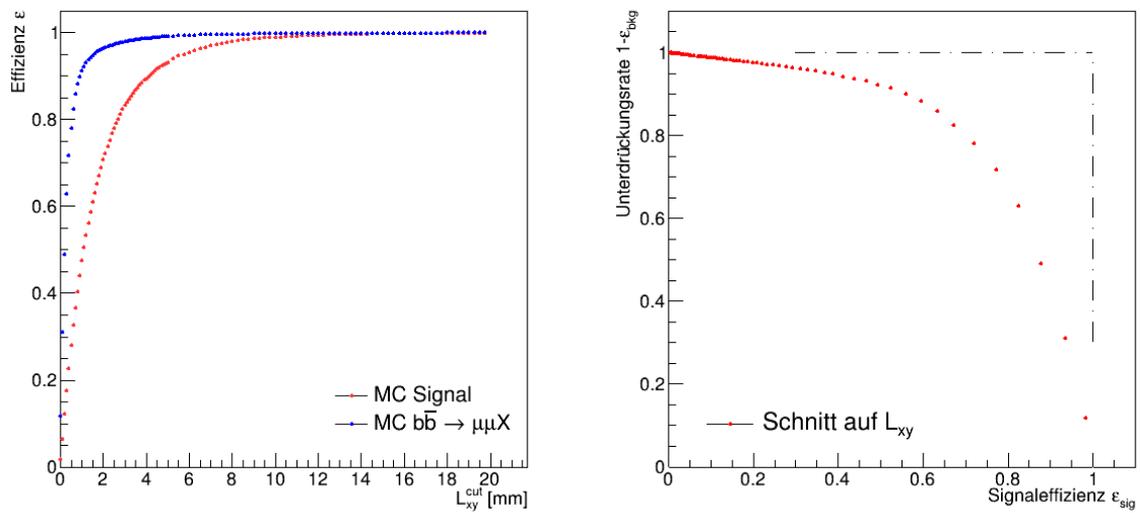


Abbildung 19: Links: Errechnete Effizienzen für Signal ε_{sig} (rot) und Untergrund ε_{bkg} (blau). Rechts: Grenzwertoptimierungskurve (ROC) für steigenden Schnitt auf L_{xy}^{cut} .

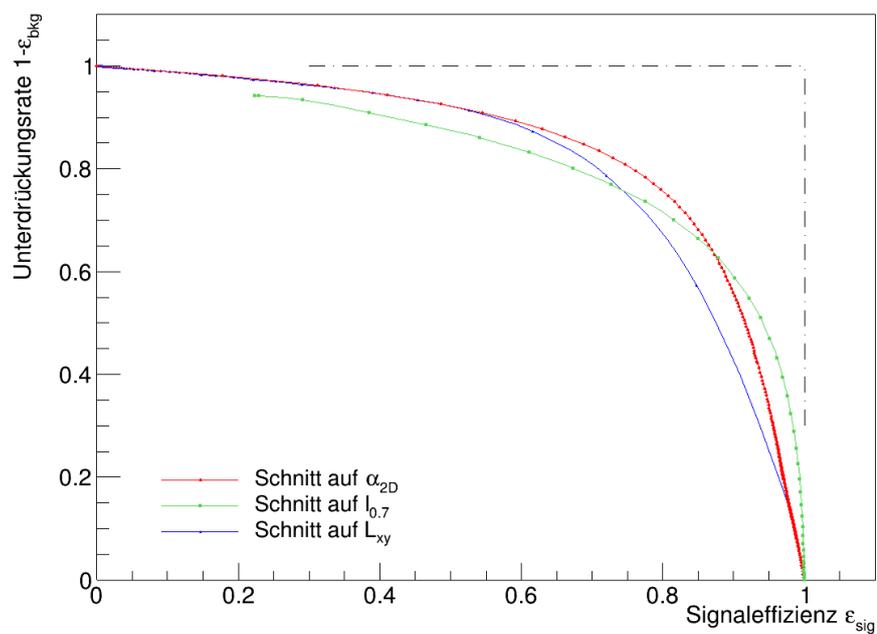


Abbildung 20: Vergleich der drei ROC-Graphen für Schnitte auf jeweils einen der drei Parameter.

ATLAS-Daten enthalten, eingelesen und wie zuvor eine Selektion anhand des dem Arbeitspunkt zugehörigen Schnittwertes vorgenommen (siehe Anhang Abb. 37 - 39). In den Abbildungen ist ein direkter Vergleich der gemessenen Massenverteilungen der invarianten $\mu^+\mu^-K^\pm$ -Massen der B^\pm -Kandidaten vor und nach dem Schnitt auf den Parameter zu sehen (Abb. 21). Es lässt sich gut erkennen, dass durch den Schnitt der Untergrund besser unterdrückt wird und die Signalspitze besser erkennbar ist. Besonders durch Schneiden auf die Isolation kann eine starke Unterdrückung der Untergrunds erzielt werden. In den Abbildungen wurden zur verbesserten Sichtbarkeit der Graphen die Punkte mit geraden Linien verbunden.

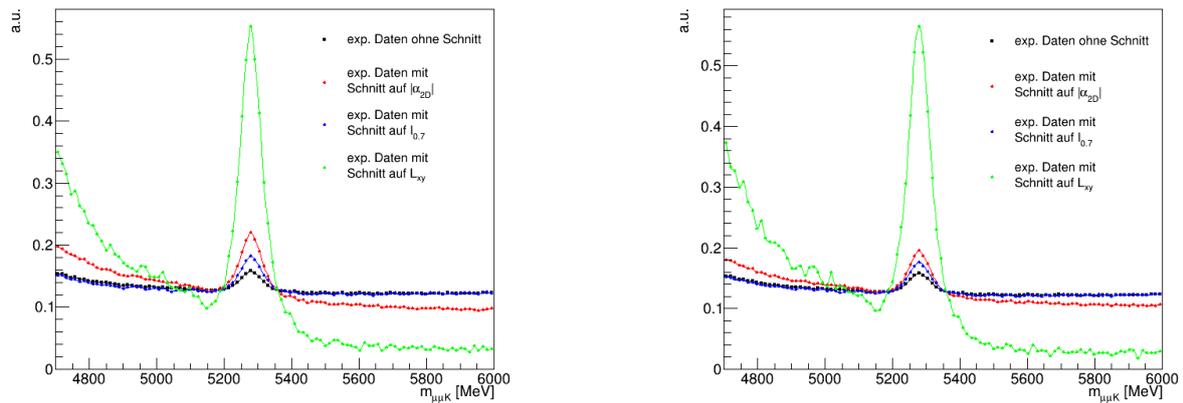


Abbildung 21: Vergleich der Schnitte auf die drei unterschiedlichen Parameter bei Anwendung auf ATLAS-Daten von 2015 und 2016 entsprechend $\varepsilon_{sig} = 0,8$ (links) und $\varepsilon_{sig} = 0,9$ (rechts) mit Hilfslinien.

5.3 Die Fisher-Diskriminante

Im nächsten Schritt soll die Fisher-Diskriminante verwendet werden, die ein lineares Diskriminanzkriterium aus mehreren Variablen bildet und effizienter den Untergrund unterdrücken und das Signal optimieren soll. Das optimiert die ROC-Kurve weitergehend dahin, dass die Kurve im Vergleich zur ROC-Kurve bei Schnitten auf eine einzelne Variable eine lokal stärkere Krümmung nahe des optimalen Punktes auf der idealen ROC-Kurve erhält und weiter in die Nähe dieses optimalen Punktes bei (1|1) verschoben wird.

5.3.1 Training auf Monte-Carlo-Daten

Die Implementierung und Verwendung dieser Methode, soll auf zwei Weisen erfolgen: Zum einen durch eine selbst implementierte Fisher-Diskriminante, die das algebraische Vorgehen aus Kapitel 3 verwendet. Zum anderen kommt eine bereits in TMVA [7] (Toolkit for Multi Variate Data Analysis) vorimplementierte Fisher-Diskriminante zum Einsatz. TMVA ist ein in ROOT [8] integriertes Paket zur Datenanalyse mittels Machine Learning und multivariater Methoden, das auch die letzte verwendete Methode, Boosted Decision Trees (Kap. 5.4), beinhaltet.

5.3.2 Eigene Implementierung der Fisher Diskriminante

Hierzu müssen zunächst die Koeffizienten der linearen Teststatistik $t(\vec{x})$ ermittelt werden, mittels welcher dann erneut die Signaleffizienz und Untergrundunterdrückung untersucht werden. Im ersten Schritt werden die arithmetischen Mittelwerte $(M_k)_i$ aller Parameter in den zwei Klassen Signal und Untergrund berechnet. Dann kann mittels dieser Vektoren von arithmetischen Mitteln der jeweiligen Parameter das arithmetische Mittel M gebildet werden. Im nächsten Schritt kann die im Kapitel 3 eingeführte Kovarianzmatrix S_W berechnet werden. Die in ROOT beinhaltete Klasse TMatrixD bietet eine Umgebung, durch die das Rechnen mit Matrizen und insbesondere das Invertieren derselben erleichtert wird. Mit der Inversen der Kovarianzmatrix S_W^{-1} kann nun der Koeffizientenvektor \vec{a} berechnet werden. Für die Koeffizienten erhält man:

$$\begin{aligned} a_{\alpha_{2D}} &= -0,244 \\ a_{I_{0.7}} &= 2,997 \\ a_{L_{xy}} &= 0,171 \end{aligned} \tag{11}$$

Die Faktoren können immer einen relativen Vorfaktor k und ein relativen Offset f haben (Kap. 3). Zur besseren Vergleichbarkeit wurden die oben gezeigten Werte mit einem Faktor von $k = 10^4$ skaliert. Dadurch erhält man folgende in Abb. 22 zu sehende Teststatistik:

$$t_{F,eigeneImpl.} = a_{\alpha_{2D}} \cdot x_{\alpha_{2D}} + a_{I_{0.7}} \cdot x_{I_{0.7}} + a_{L_{xy}} \cdot x_{L_{xy}} \tag{12}$$

Die Statistik zeigt, aufgeteilt in die zwei Klassen, eine bessere Separabilität von Signal und Untergrund durch einen geeigneten Schnittwert auf $t_{F,eigeneImpl.}(x)$ als die Separabilität durch Schnitte auf einzelne Variablen. Die Effizienzgraphen zeigen ein kleines Intervall, in denen die Effizienzen von 1 auf 0 abfallen, wenn der Schnittwert $t_{F,eigeneImpl.}^{cut}$ verändert wird (Abb. 23). Das führt im ROC-Graphen dazu, dass die Kurve sich mehr der einer idealen Grenzwertoptimierungskurve annähert.

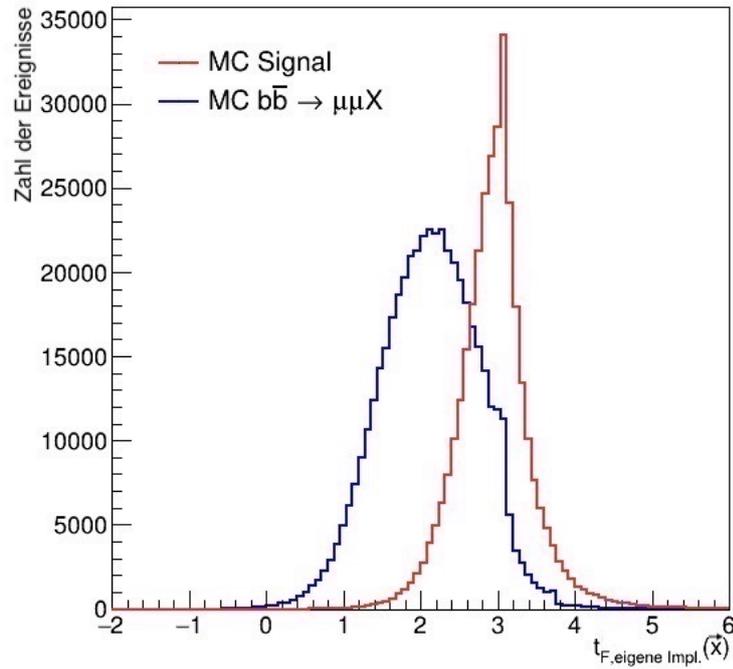


Abbildung 22: Die Teststatik $t_{F,eigeneImpl.}(x)$ für Untergrund und Signal auf Basis des MC-Testdatensatzes für Signal bzw. Untergrund.

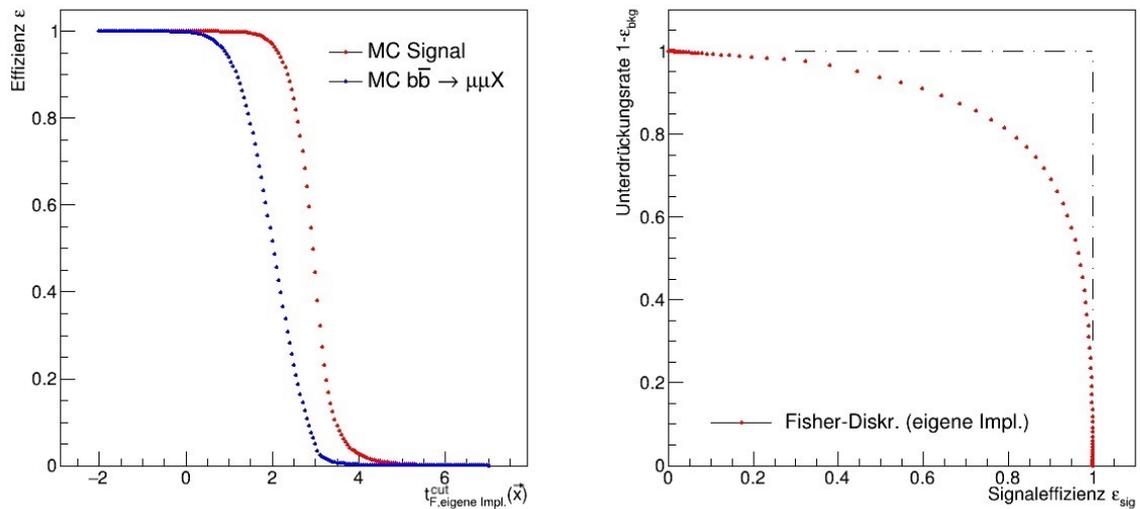


Abbildung 23: Links: Errechnete Effizienzen für Signal ε_{sig} (rot) und Untergrund ε_{bkg} (blau). Rechts: Grenzwertoptimierungskurve (ROC) für einen Schnitt auf $t_{F,eigeneImpl.}(x)$.

5.3.3 TMVA: Fisher Diskriminante

Nach der eigenen Implementierung und dem Trainieren der Fisher-Diskriminante wurde nun die in TMVA bereits implementierte Fisher-Diskriminante verwendet. Dabei lädt TMVA jeweils ein Trainings- und Testpaar von Signal und Untergrund NTupeln. Die Ausgabe dieser Methode besteht ebenfalls aus den Koeffizienten, sowie einem relativen Offset f , der allerdings dazu dient, die Teststatistik um den Ursprung zu zentrieren (Abb. 24).

$$\begin{aligned}
 a_{\alpha_{2D},TMVA} &= -0,188 \\
 a_{I_{0.7},TMVA} &= 2,167 \\
 a_{L_{xy},TMVA} &= 0,112 \\
 f &= -1,760
 \end{aligned} \tag{13}$$

Die daraus resultierende Teststatistik wird für die Berechnung des Diskriminanzwertes $t_{F,TMVA}$ der ATLAS-Daten verwendet, auf den dann geschnitten wird. Genau wie zuvor zeigt die Statistik eine verbesserte Trennung von Signal und Untergrund in den Parametern (Abb. 24) im Vergleich zur Trennung durch eindimensionale Schnitte auf einzelne Parameter. Die ROC-Kurve und Effizienzgraphen haben die selben Charakteristiken, wie bei der eigenen Implementierung (Abb. 25).

$$t_{F,TMVA} = a_{\alpha_{2D},TMVA} \cdot x_{\alpha_{2D}} + a_{I_{0.7},TMVA} \cdot x_{I_{0.7}} + a_{L_{xy},TMVA} \cdot x_{L_{xy}} + f \tag{14}$$

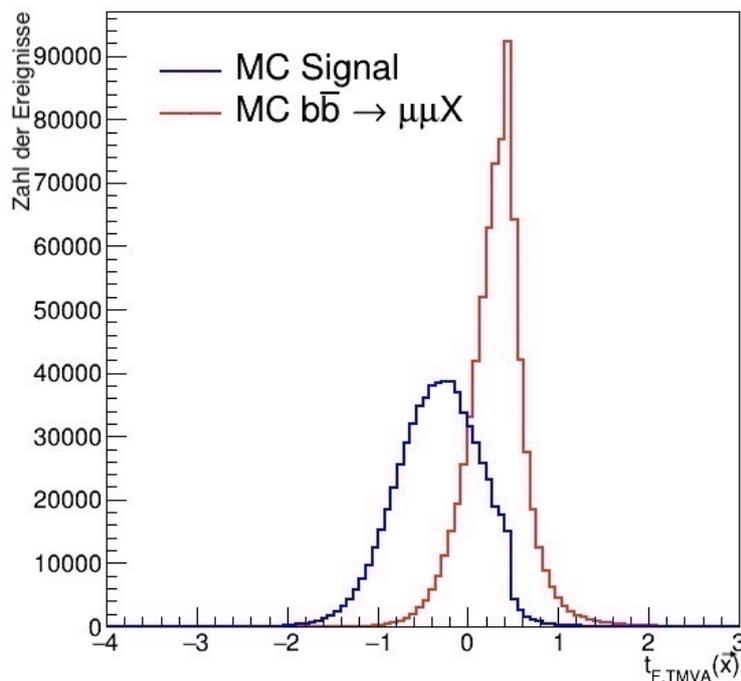


Abbildung 24: Die Teststatistik $t_{F,TMVA}(x)$ für Untergrund und Signal auf Basis des MC-Testdatensatzes für Signal bzw. Untergrund.

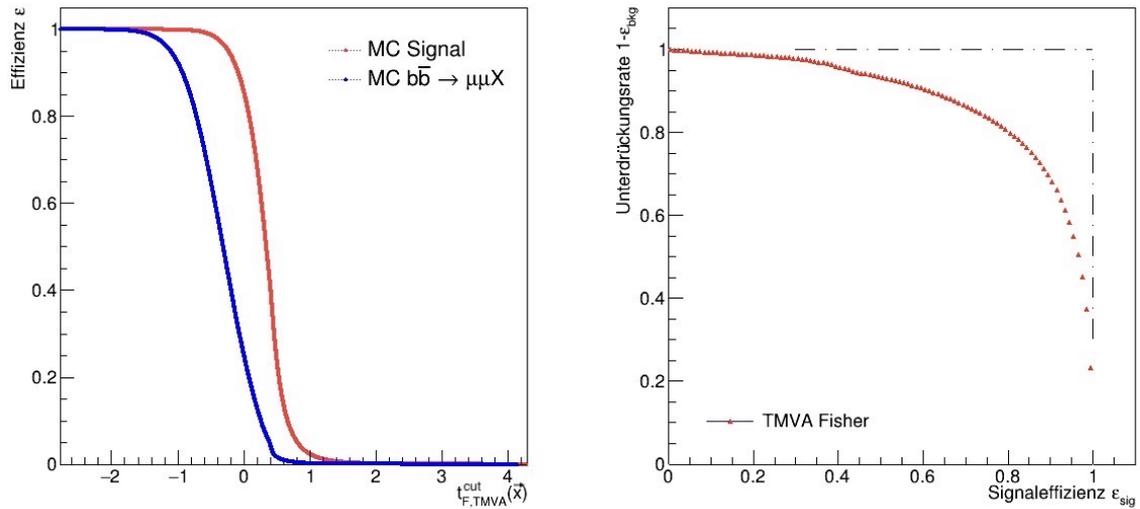


Abbildung 25: Links: Errechnete Effizienzen für Signal ε_{sig} (rot) und Untergrund ε_{bkg} (blau). Rechts: Grenzwertoptimierungskurve (ROC) (rechts) für einen Schnitt auf $t_{F, TMVA}(x)$.

Vergleich beider Implementierungen

Ein direkter Vergleich der beiden Fisher-Diskriminantenimplementierungen zeigt eindeutige Unterschiede in der Form der Statistik (Abb. 26). Ein detaillierter Vergleich zeigt auch, dass die Koeffizienten keinen konsistenten Proportionalitätsfaktor haben (Gl. 15) und zudem einen Offset von $f = 1,760$.

$$\begin{aligned}
 \frac{a_{\alpha_{2D}}}{a_{\alpha_{2D}, TMVA}} &= 1,298 \\
 \frac{a_{I_{0.7}}}{a_{I_{0.7}, TMVA}} &= 1,383 \\
 \frac{a_{L_{xy}}}{a_{L_{xy}, TMVA}} &= 1,527
 \end{aligned} \tag{15}$$

Es findet sich kein gemeinsamer Faktor, jedoch sind die ROC-Kurven der beiden Implementierungen identisch (Abb. 26 rechts). Dies ist ein Zeichen für die Gleichwertigkeit beider Implementierungen.

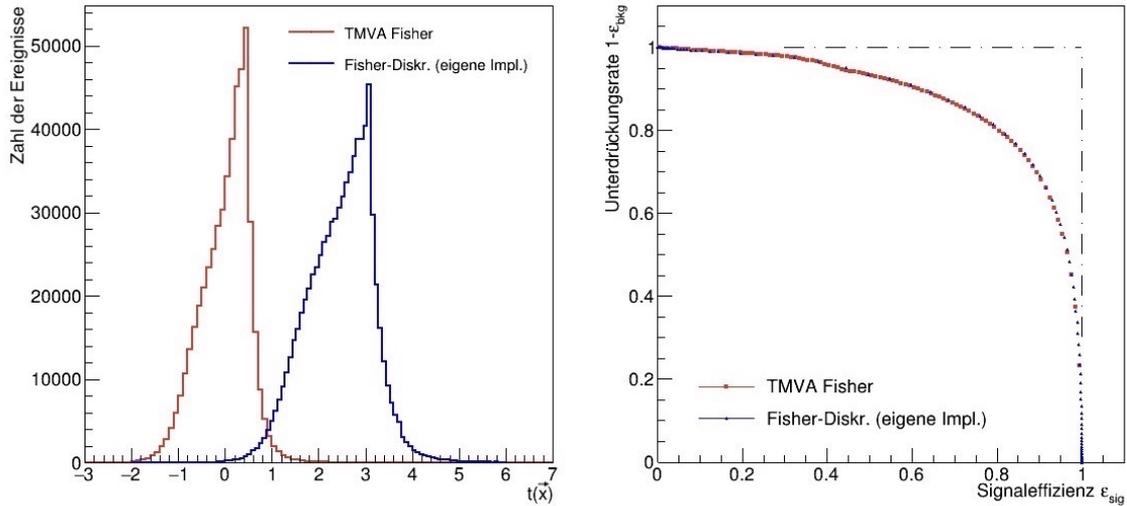


Abbildung 26: Links: Vergleich der Teststatistiken der beiden Implementierungen der Fisher-Diskriminante durch eine eigene Implementierung (blau) und TMVA (rot). Rechts: Vergleich der ROC-Graphen der beiden Implementierungen der Fisher-Diskriminante durch eine eigene Implementierung (blau) und TMVA (rot).

5.3.4 Anwendung auf reale Messdaten

Nach dem Training der beiden Implementierungen der Fisher-Diskriminante wurden beide Teststatistiken auf den ATLAS-Datensatz von 2015 und 2016 angewendet. Dafür werden aus den NTupeln der Messdaten die Werte für die Parameter $|\alpha_{2D}|$, $I_{0.7}$ und L_{xy} eingelesen und anschließend werden mittels der Koeffizienten der jeweiligen Methoden die Werte $t_{F,\text{eigeneImpl.}}$ und $t_{F,TMVA}$ für jedes Ereignis berechnet. Im Anschluss werden in den ROC-Graphen der selbst-implementierten Fisher-Diskriminante und der in TMVA implementierten Diskriminanzfunktion zwei Arbeitspunkte ausgewählt: Bei einer Signaleffizienz ϵ_{sig} von 80 Prozent und 90 Prozent. In den Effizienzgraphen beider Implementierungen können danach die Schnittwerte $t_{F,\text{eigeneImpl.}}^{\text{cut}}$ bzw. $t_{F,TMVA}^{\text{cut}}$ ermittelt werden.

Nun wird $m_{\mu\mu K}$ aus den ATLAS-Daten in die Histogramme eingefüllt (Abb. 27, 28), die den Wert für $t_{F,\text{eigeneImpl.}}^{\text{cut}}$ und $t_{F,TMVA}^{\text{cut}}$ jeweils überschreiten. Die Histogramme sind bezüglich ihrer Fläche normiert, das bedeutet das die Fläche unterhalb der Kurve gleich normiert wird, um die beiden Histogramme vergleichen zu können. In den Bildern ist zu sehen, dass der Untergrund effektiv unterdrückt wird, aber die Untergrundunterdrückung bei einer Signaleffizienz von ϵ_{sig} 80 Prozent nur marginal größer ist als die bei 90 prozentiger Signaleffizienz (Abb. 27).

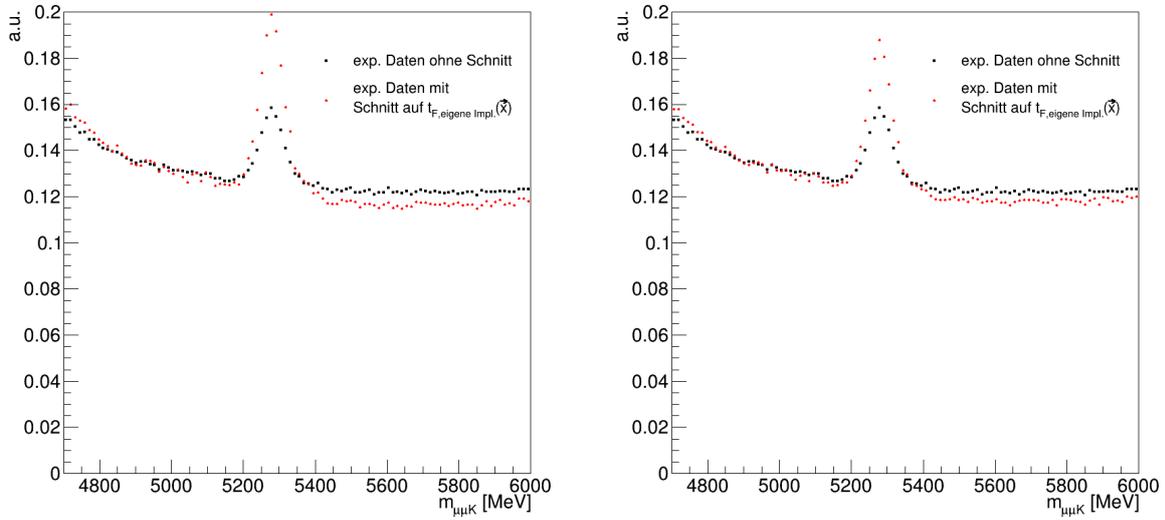


Abbildung 27: ATLAS-Daten von 2015 und 2016 mit Schnitt auf den berechneten Diskriminanzwert $t_{F,eigeneImpl}$. entsprechend $\epsilon_{sig} = 0,8$ (links) und $\epsilon_{sig} = 0,9$ (rechts). Die beiden Histogramme sind jeweils auf die gleiche Fläche normiert.

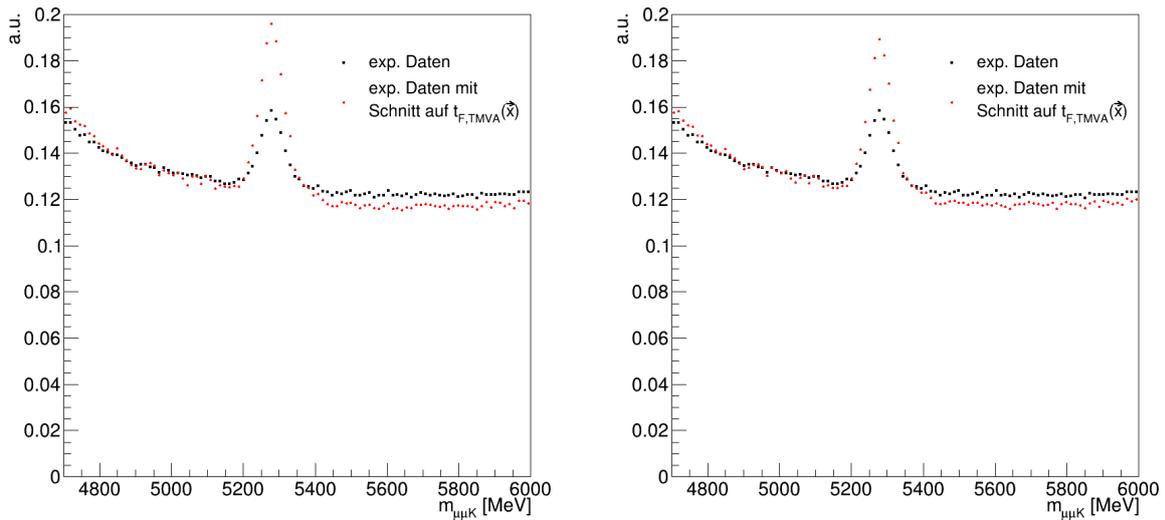


Abbildung 28: ATLAS-Daten von 2015 und 2016 mit Schnitt auf den berechneten Diskriminanzwert $t_{F,TMVA}$ entsprechend $\epsilon_{sig} = 0,8$ (links) und $\epsilon_{sig} = 0,9$ (rechts).

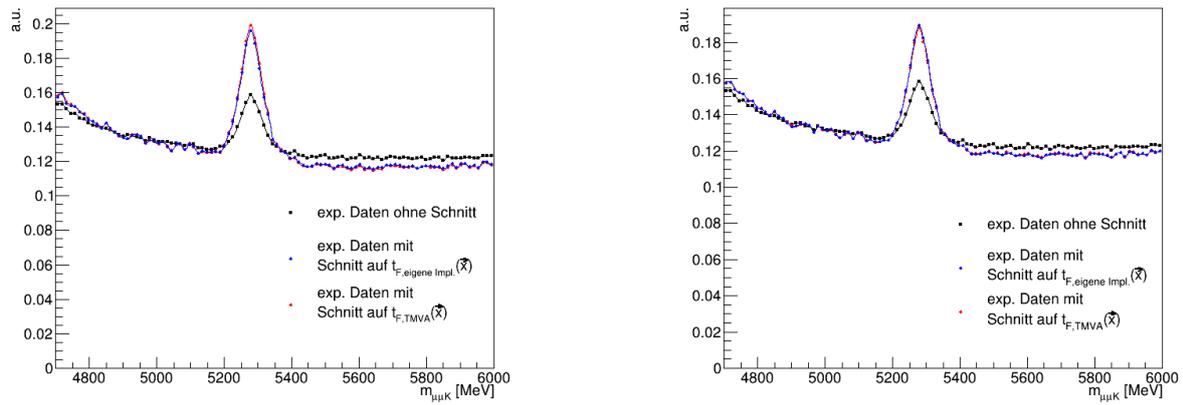


Abbildung 29: Vergleich der beiden Implementierungen der Fisherdiskriminante bei Anwendung auf ATLAS-Daten ($1,792 \cdot 10^7$ Ereignisse) von 2015 und 2016 mit Schnitt auf den berechneten Diskriminanzwert $t_{F,eigeneImpl.}$ ($5,885 \cdot 10^6$ Ereignisse) und $t_{F, TMVA}$ ($5,887 \cdot 10^6$ Ereignisse) für $\varepsilon_{sig} = 0,8$ (links) und $\varepsilon_{sig} = 0,9$ (rechts) mit Hilfslinien.

5.4 Boosted Decision Trees

Im letzten Schritt sollen Boosted Decision Trees zum Einsatz kommen, die den Untergrund optimaler unterdrücken sollen. Die hierfür genutzte Klasse ist in ROOT bereits implementiert. Diese Methode verwendet `LTMVA::MethodBTD`, das in Kapitel 3.3 eingeführte Verfahren.

5.4.1 Training auf Monte-Carlo-Daten

Wie bereits bei der Fisher-Diskriminante in TMVA werden hier wieder vier Datensätze eingelesen. Zunächst wird ein Modell für den Boosted Decision Tree mit dem Trainingsdatensatz erstellt und die Gewichte der einzelnen Knoten gespeichert [8]. Dabei wurden die folgenden Einstellungen vorgenommen (Tab. 2):

Option	Wert	Funktion
NTrees	850	Zahl der unabhängig trainierten Decision Trees
MinNodeSize	2,5%	relative Größe der Knoten zur Gesamtheit
MaxDepth	3	maximale Tiefe bevor das weitere Spalten der Knoten endet
BoostType	AdaBoost	Art des Boostalgorithmus
AdaBoostBeta	0,5	Lernrate des Boostalgorithmus
SeparationType	GiniIndex	Separationkriterium für das Spalten der Knoten im folgenden Schritt
nCuts	20	Anzahl an Intervallpunkten zur Ermittlung eines optimalen Schnitts für das Spalten der Knoten im folgenden Schritt

Tabelle 2: Die für die Implementierung des BDTs getroffenen Einstellungen und ihre Bedeutungen [8].

Die Zahl der durchlaufenen BDT NTrees ist im Zusammenwirken mit den anderen Optionen wichtig für das optimale Training des BDT, da das erstellte Modell mit den eingelesenen Datensätzen übertrainiert werden kann. Ist dies der Fall, sinkt die Effektivität des Modells hinsichtlich der Untergrundunterdrückung. Gleiches gilt für untertrainierte BDT, da dann die entwickelten Kriterien zu viele Signalereignisse aussortieren. Nach Anwendung der BDT Implementierung in TMVA erhält man wie schon bei Fisher eine Statistik mit dem zugehörigen Diskriminanzwert $t_{BDT}(x)$ (Abb. 30).

Nach Beendigung des Trainings werden die Gewichte, die das errechnete Modell beschreiben, in eine entsprechende Datei geschrieben. Diese wird für die Anwendung auf die ATLAS-Daten benötigt. Auch hier lässt sich in Abhängigkeit des Schnitts t_{BDT}^{cut} ein ROC-Graph plotten (Abb. 31).

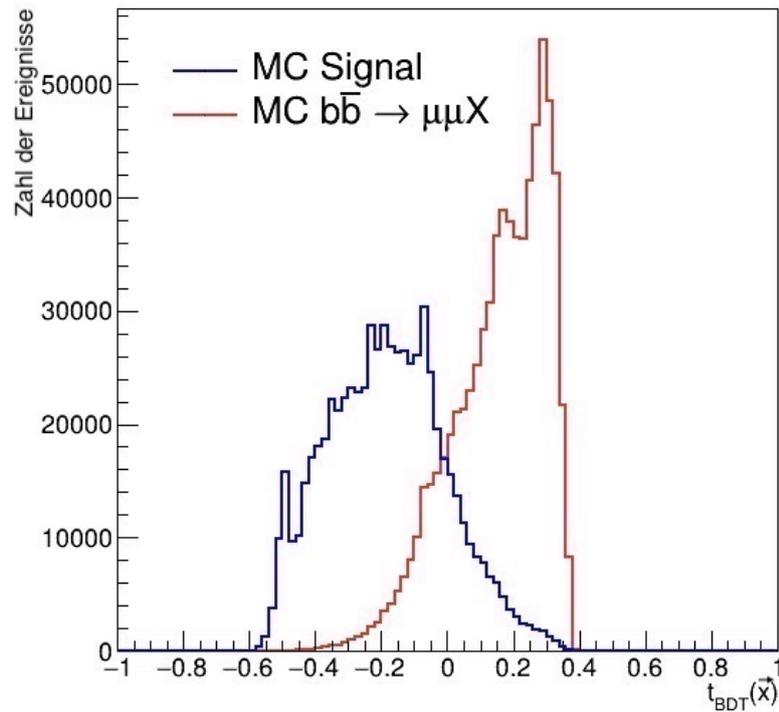


Abbildung 30: Die Teststatik $t_{BDT}(x)$ für Untergrund und Signal auf Basis der entsprechenden MC-Testdatensätze.

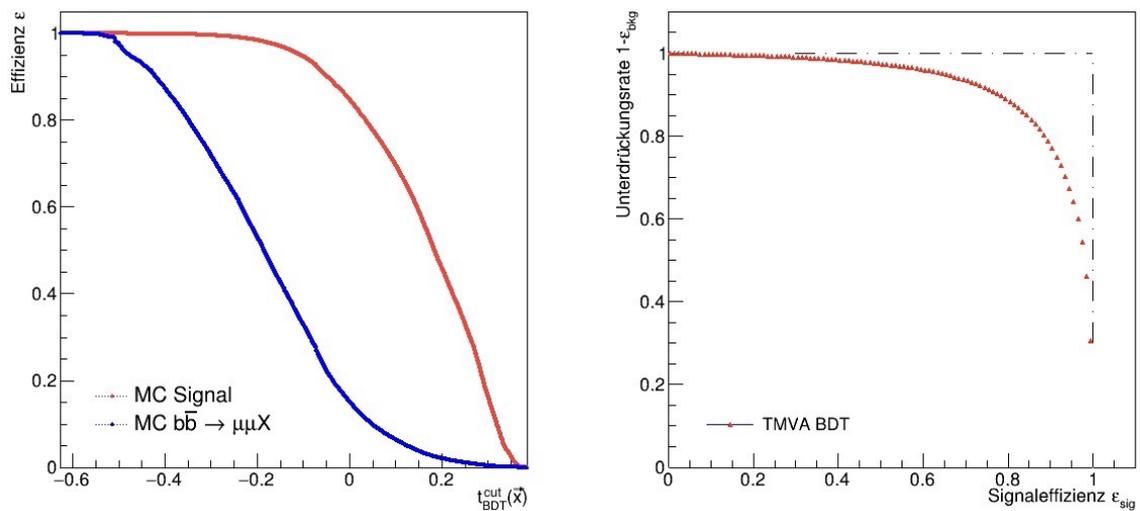


Abbildung 31: Links: Errechnete Effizienzen für Signal ϵ_{sig} (rot) und Untergrund ϵ_{bkg} (blau). Rechts: Grenzwertoptimierungskurve (ROC) (rechts) für einen Schnitt auf $t_{BDT}(x)$.

5.4.2 Anwendung auf reale Messdaten

Das gefundene Modell wird nun auf die Messdaten angewendet. Die im BDT-Training entwickelten Gewichte werden auf die gemessenen Ereignisse angewendet und ergeben den Wert der Teststatistik t_{BDT} . Damit wird die Teststatistik $t_{BDT}(x)$ für die Messdaten ermittelt auf welche im Anschluss das gleiche Verfahren angewendet wird wie bei den Fisher-Diskriminanten. Dem Effizienzgraphen werden die Schnitte t_{BDT}^{cut} von zwei gewählten Arbeitspunkten auf der ROC-Kurve mit jeweils $\varepsilon_{sig} = 0,8$ bzw. $\varepsilon_{sig} = 0,9$ entnommen (Abb. 31). Dadurch kann eine deutliche Unterdrückung des Untergrundes in den Messdaten erreicht werden (Abb. 32).

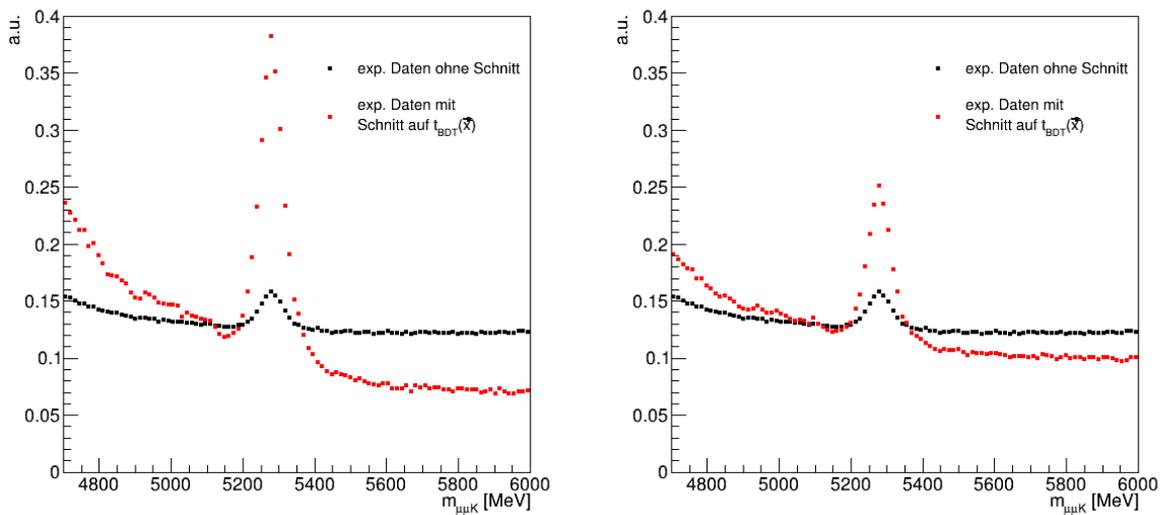


Abbildung 32: ATLAS-Daten von 2015 und 2016 mit Schnitt auf den berechneten Diskriminanzwert $t_{F,BDT}$ entsprechend $\varepsilon_{sig} = 0,8$ ($1,726 \cdot 10^6$ Ereignisse) (links) und $\varepsilon_{sig} = 0,9$ ($3,992 \cdot 10^6$ Ereignisse) (rechts).

6 Vergleich der Methoden

Zum Ende dieser Analyse sollen nun die drei im vorherigen Kapitel beschriebenen Methoden in ihrer Effektivität und Eignung zur Unterdrückung des Untergrundes verglichen werden. Dazu können die ROC-Kurven verwendet werden, da sie die gleiche Grundgesamtheit, also die gleichen Datensätze zum Trainieren und Testen, verwenden.

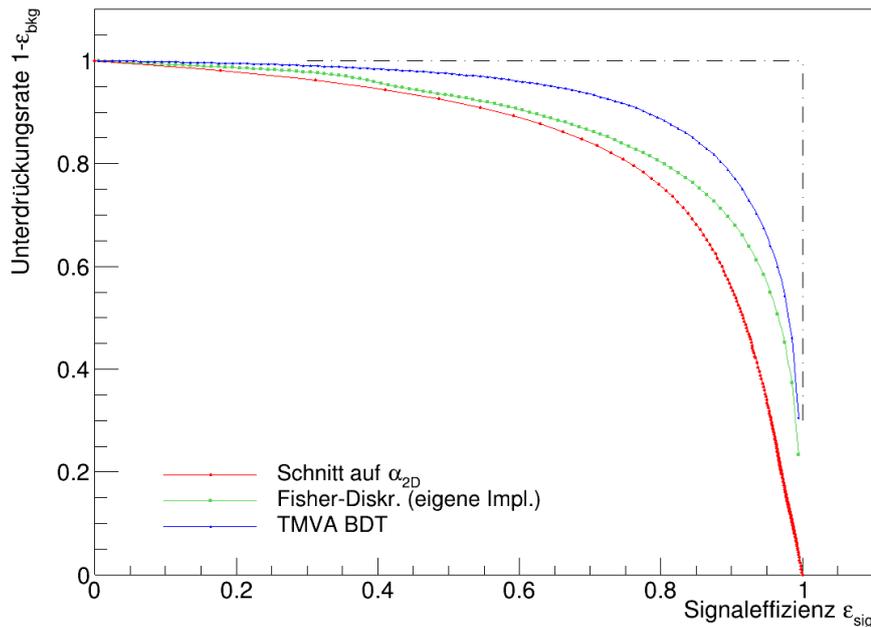


Abbildung 33: Vergleich der drei ROC-Graphen der drei Methoden. Für die einfachen Schnitte wurde hier der ROC-Graph von $|\alpha_{2D}|$ verwendet, da dieser die optimale Kurve besitzt im Vergleich zu den anderen beiden Parametern.

In Abb. 33 wird sichtbar, dass die Fisher-Methode die Effizienz des Signals unter Minimierung des Untergrundes verbessert. Es handelt sich hier um die einfachste Form einer multivariaten Methode. Die BDT-Methode ist am effektivsten, da bei gleicher Signalstärke die beste die Unterdrückung des Untergrundes erzielt wird.

Bei Betrachtung der $m_{\mu\mu K}$ -Verteilung in den Messdaten im Vergleich zu den Messdaten nach Anwendung der Schnittwerte einzelner Methoden ist zu erkennen, dass BDT sichtbar die optimale Unterdrückung erreicht (Abb. 34), wengleich der Unterschied zwischen den Messdatenpunkten nach Fisher und Schnitten auf $|\alpha_{2D}|$ im Vergleich zu BDT klein zu sein scheint. Der Vergleich der beiden letzteren Methoden zeigt aber deutlich eine Verbesserung der Effektivität bei Verwendung multivariater Methoden mit Maschinellem Lernen (Abb. 33).

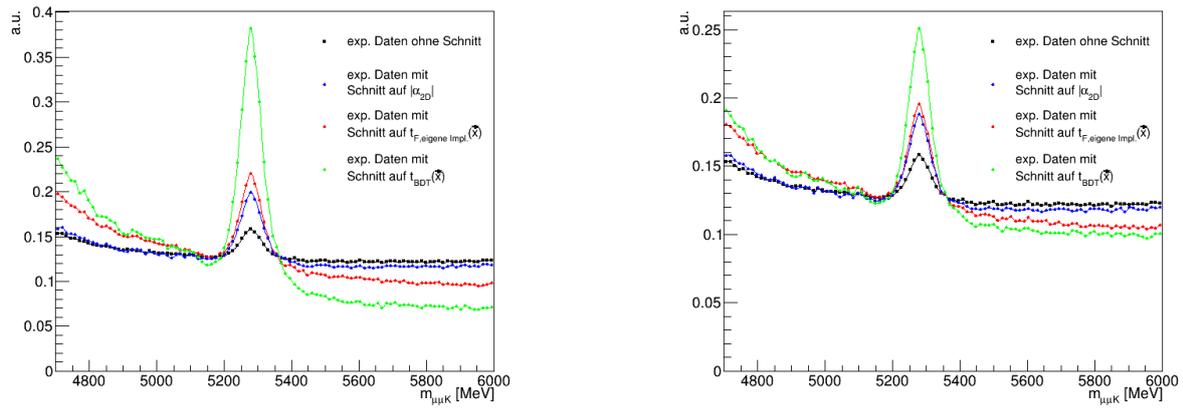


Abbildung 34: Vergleich der verwendeten Methoden bei Anwendung auf ATLAS-Daten von 2015 und 2016 mit Schnitt auf $|\alpha_{2D}|$, $t_{F,eigeneImpl}^{cut}$ bzw. t_{BDT}^{cut} , entsprechend $\varepsilon_{sig} = 0,8$ (links) und $\varepsilon_{sig} = 0,9$ (rechts), mit Hilfslinien.

7 Zusammenfassung und Ausblick

In dieser Arbeit sollten mono- und multivariate Methoden hinsichtlich ihrer Effektivität zur Signaloptimierung bei bestmöglicher Untergrundunterdrückung verglichen werden.

Dazu wurden zunächst einfache Schnitte je auf einen von drei Parametern gesetzt, um Untergrundereignisse zu entfernen. Dafür wurde je ein MC-Datensatz von Signal- und ein MC-Datensatz von Untergrundereignissen durch die Betrachtung der Signaleffizienz und Untergrundunterdrückungsrate analysiert, um im Anschluss zwei Arbeitspunkte mit einer Signaleffizienz von 80 bzw. 90 Prozent zu wählen. Der MC-Signaldatensatz enthält rekonstruierte B^+ -Kandidaten für den Zerfall $B^+ \rightarrow J/\psi K^+$ mit $J/\psi \rightarrow \mu^+ \mu^-$. Der MC-Untergrunddatensatz enthält die Zerfälle der Form $b\bar{b} \rightarrow \mu^+ \mu^- X$ ohne $B^\pm \rightarrow J/\psi K^\pm$. Durch die Arbeitspunkte lässt sich ein zugehöriger Schnitt auf einen der drei Parameter finden, auf den dann in den ATLAS-Daten geschnitten werden kann. Dadurch wird der Untergrund unterdrückt, aber es geht unvermeidlich Signal verloren. Hier wurde auf die Parameter $|\alpha_{2D}|$, $I_{0.7}$ und L_{xy} geschnitten.

Im nächsten Schritt kam eine lineare Methode zum Berechnen einer Teststatistik zum Einsatz — die Fisher-Diskriminante. Dabei wurden zwei Implementierungen verwendet. Zum einen wurden die Koeffizienten der linearen Statistik mittels der algebraischen Methode nach Fisher selbst berechnet (Kap. 3). Zum anderen wurden diese mit den berechneten Koeffizienten der in TMVA bereits implementierten Fisher-Diskriminante verglichen. Die ROC-Graphen der beiden Implementierungen sind identisch, trotz der Unterschiede der beiden Teststatistiken aufgrund eines inkonsistenten Proportionalitätsfaktors k und relativen Offsets $f = 1,760$. Beide Teststatistiken wurden auf ATLAS-Daten aus den Jahren 2015 und 2016, in denen die Zerfälle $B^\pm \rightarrow J/\psi K^\pm$ mit $J/\psi \rightarrow \mu^+ \mu^-$ rekonstruiert wurden, angewendet und ihre Verteilungen $t_{F,eigeneImpl.}(x)$ und $t_{F,TMVA}(x)$ für die einzelnen Ereignisse berechnet.

Zuletzt wurden Boosted Decision Trees mit dem Trainingsdatensatz verwendet, um ein Modell mit Gewichten zu erstellen, die den Untergrund optimal unterdrücken und das Signal erhalten können sollen. Die gefundenen Gewichte wurden auf dieselben ATLAS-Daten nach vorheriger Validation mit dem Trainingsdatensatz, also der zweiten Hälfte der jeweiligen MC-Datensätze, angewendet und für jedes Ereignis der Teststatistikwert t_{BDT} berechnet.

Nachdem für die beiden letzteren Methoden die Statistiken auf den ATLAS-Daten berechnet wurden, wurde die invariante Masse $m_{\mu\mu K}$ nach Anwendung der jeweiligen Schnitte auf einen Parameter oder die Teststatistiken in Histogramme gefüllt.

Im Vergleich der Methoden ist eine verbessertes Verhältnis von Signal zu Untergrund bzw. eine verbesserte Untergrundunterdrückung bei zeitgleicher Signaloptimierung bei allen Methoden im Vergleich zu den ATLAS-Daten ohne Schnitt auf eine Variable oder Teststatistik festzustellen. Besonders deutlich wird dies in den ROC-Graphen (Abb. 33) und den Massenverteilungen $m_{\mu\mu K}$ (Abb. 34).

Zusammenfassend komme ich im Hinblick auf die anfänglich gestellte Frage zu dem Schluss, dass sich mittels multivariater Methoden auf Basis Maschinellen Lernens, den verwendeten Boosted Decision Trees, eine wesentliche Steigerung der Untergrundunterdrückung bei zeitgleicher Signalsoptimierung erreichen lässt. Durch effektives Training ohne Über- bzw. Untertrainieren auf möglichst präzise simulierten Ereignissen lässt sich der Untergrund optimal unterdrücken und das Signal optimieren.

8 Anhang

Effizienzgraphen

Ein erster Ansatz zur Bestimmung eines optimalen Arbeitspunktes bestand darin, die Differenz der beiden Effizienzen zu maximieren, um dann ein Schnittpunkt für die einzelnen Parameter zu bestimmen (Abb. 35, 36). Des Weiteren wurde noch die Signifikanz berechnet, um ebenfalls einen optimalen Schnittpunkt zu wählen. Sie errechnet sich mittels der Anzahl der Signal- bzw. Untergründereignisse $N_{sig, cut}$ bei Erfüllung des Schnittkriteriums:

$$\sigma_{sig} = \frac{N_{sig, cut}}{\sqrt{N_{bkg, cut}}} \quad (16)$$

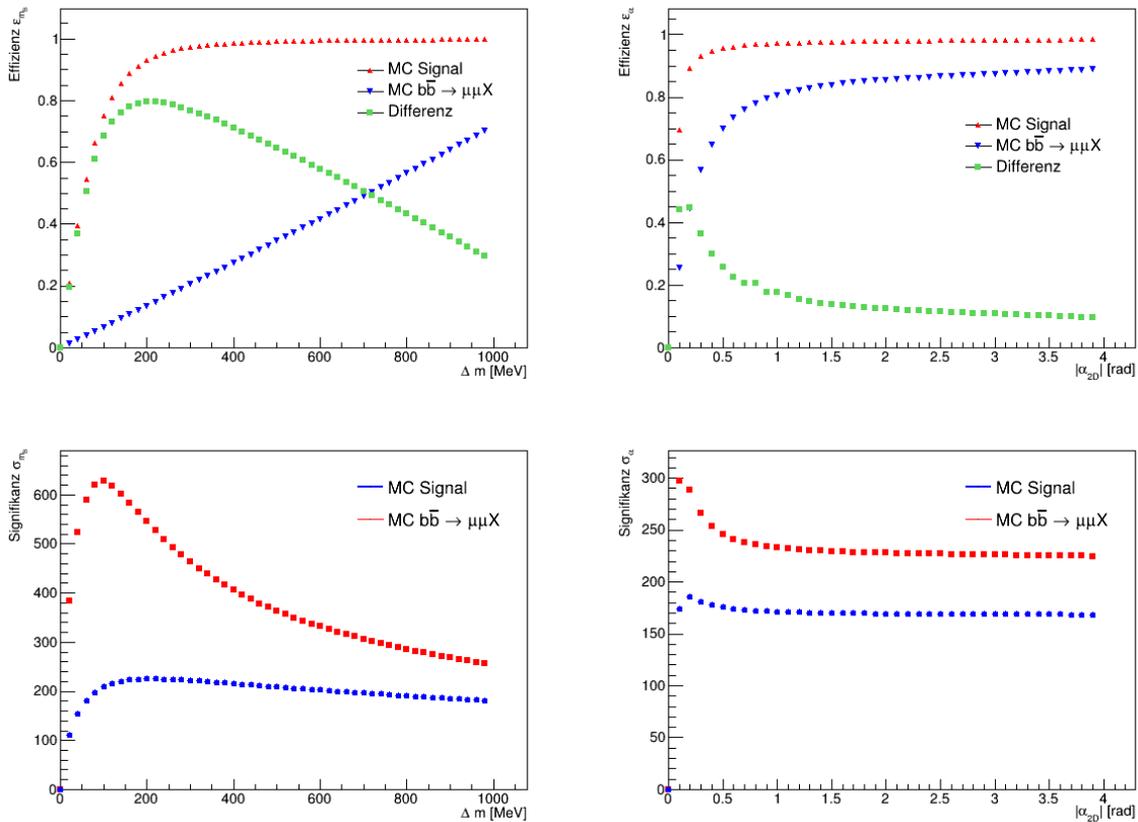


Abbildung 35: Signaleffizienz und Untergründeffizienz des Projektionswinkels $|\alpha_{2D}|$ (rechts oben), sowie seiner Signifikanz für Signal und Untergrund (rechts unten). Der Graph dient der Bestimmung eines optimalen Schnittwertes für α_{2D} . Links sind zudem die Effizienzen und Signifikanzen der Masse in Abhängigkeit des betrachteten Massenintervalls Δm symmetrisch um die Signalspitze der Massenverteilung des B -Mesons.

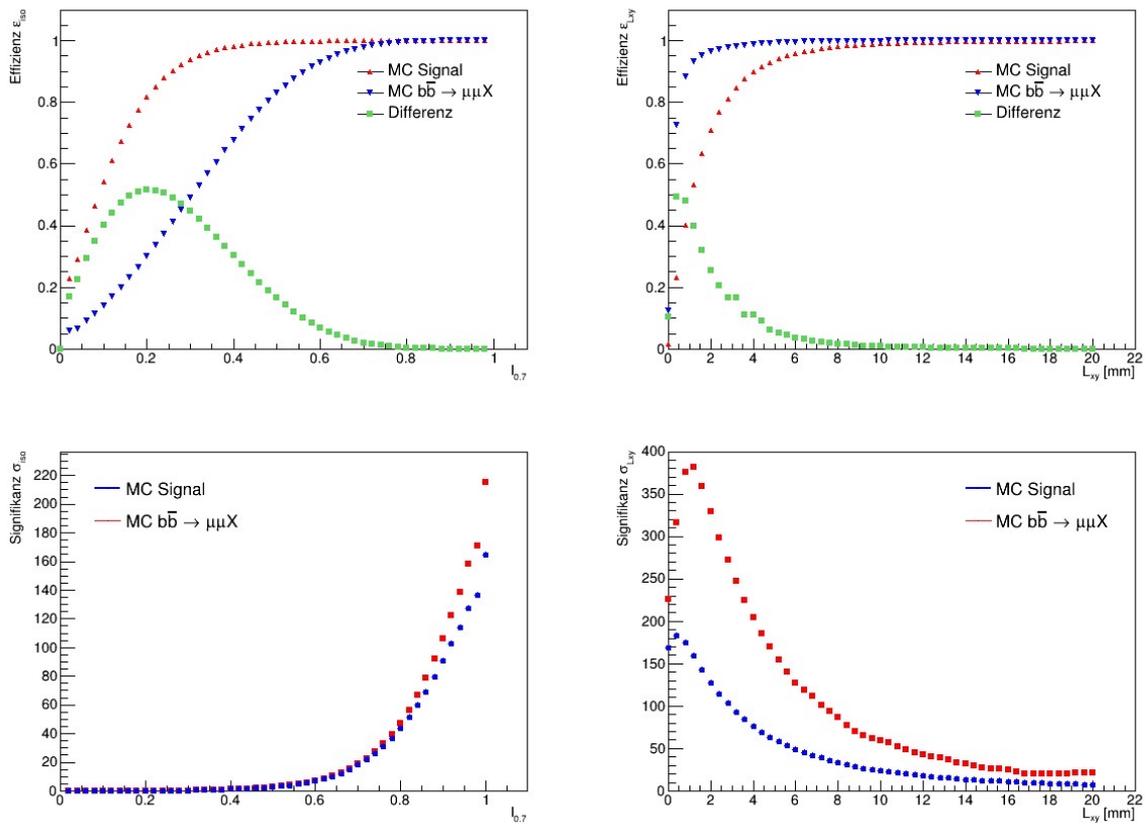


Abbildung 36: Signaleffizienz und Untergrundeffizienz der Isolation $I_{0.7}$ (oben links) und der transversalen Zerfallslänge L_{xy} (oben rechts), sowie deren Signifikanz für Signal und Untergrund (unten). Beide Graphen dienen der Bestimmung eines optimalen Schnittwertes für $I_{0.7}$ und L_{xy} .

Anwendung auf ATLAS-Daten

Die Anwendung auf reale ATLAS-Daten war ein wichtiger Teil bei der Untersuchung der Methoden zur Signalloptimierung. In Abb. 37 bis 39 sind die Schnitte auf die jeweiligen Parameter für Signaleffizienzen von 80 Prozent und 90 Prozent angewendet auf die Massenverteilungen $m_{\mu\mu K}$ zu sehen.

Auch der Vergleich der einfachen Schnitte und der Fisher-Diskriminante ist deutlich in Abb. 40 zu sehen, während in Abb. 41 der Vergleich der multivariaten Methoden entsprechend der Signaleffizienzen von 0,8 bzw. 0,9 zu sehen ist.

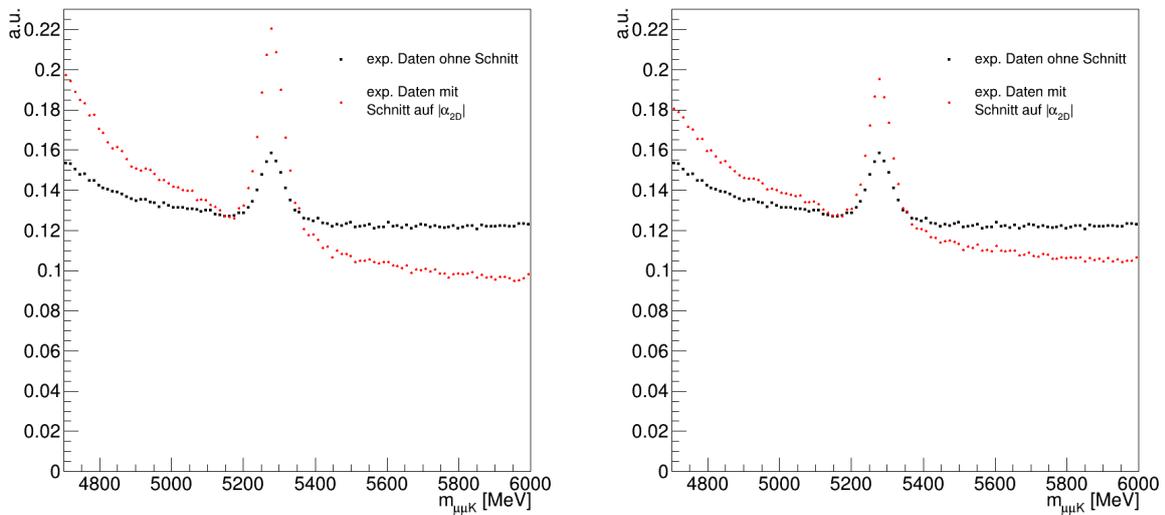


Abbildung 37: ATLAS-Daten von 2015 und 2016 mit Schnitt auf $|\alpha_{2D}|$ entsprechend $\varepsilon_{sig} = 0,8$ ($5,413 \cdot 10^6$ Ereignisse) (links) und $\varepsilon_{sig} = 0,9$ ($7,733 \cdot 10^6$ Ereignisse) (rechts).

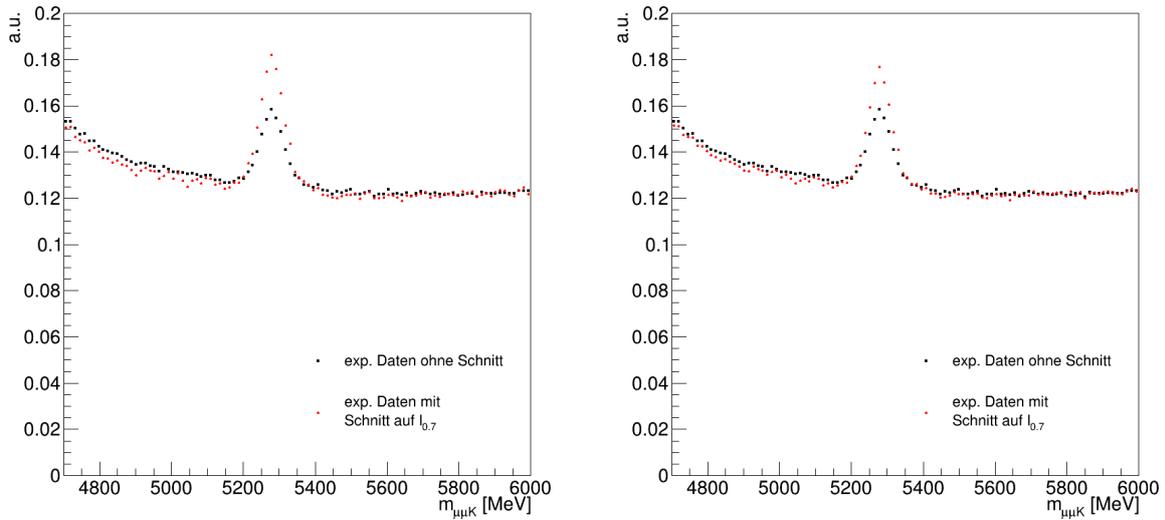


Abbildung 38: ATLAS-Daten von 2015 und 2016 mit Schnitt auf $I_{0.7}$ entsprechend $\epsilon_{sig} = 0,8$ ($7,316 \cdot 10^6$ Ereignisse) (links) und $\epsilon_{sig} = 0,9$ ($9,265 \cdot 10^6$ Ereignisse) (rechts).

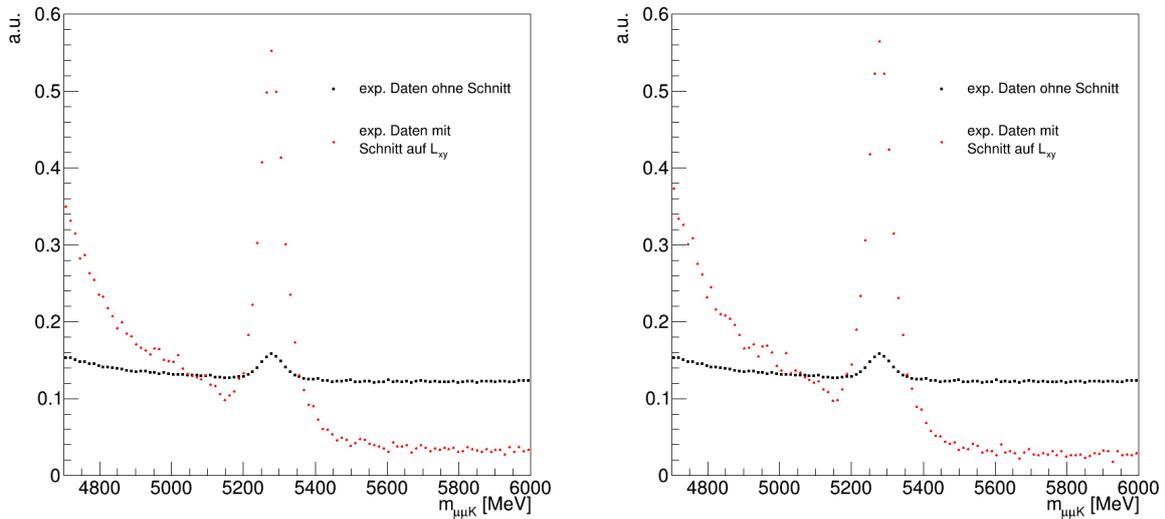


Abbildung 39: ATLAS-Daten von 2015 und 2016 mit Schnitt auf L_{xy} entsprechend $\epsilon_{sig} = 0,8$ ($3,539 \cdot 10^5$ Ereignisse) (links) und $\epsilon_{sig} = 0,9$ ($2,021 \cdot 10^5$ Ereignisse) (rechts).

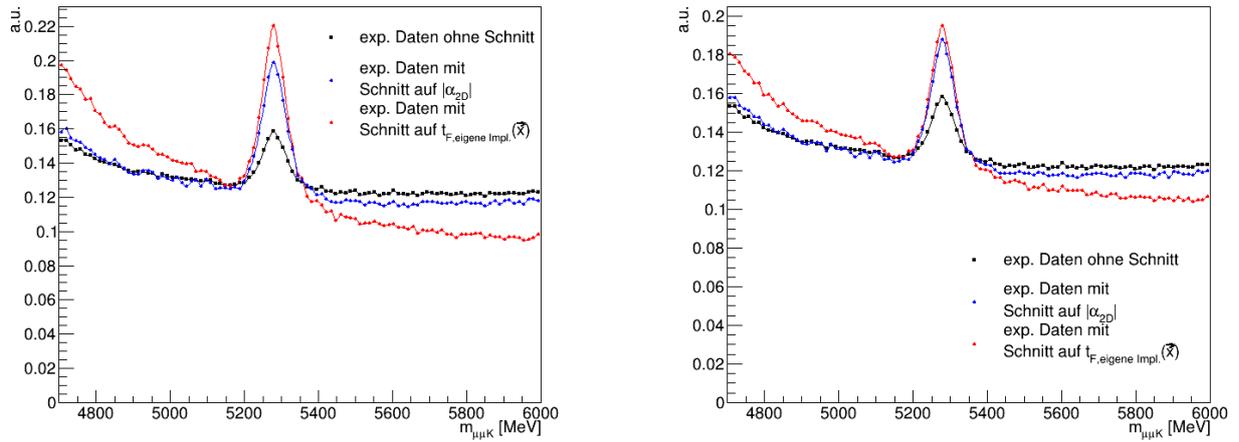


Abbildung 40: Vergleich der selbst implementierten Fisher-Diskriminante mit den einfachen Schnitten auf jeweils α_{2D} , $I_{0.7}$ und L_{xy} bei Anwendung auf ATLAS-Daten von 2015 und 2016 entsprechend $\varepsilon_{sig} = 0,8$.

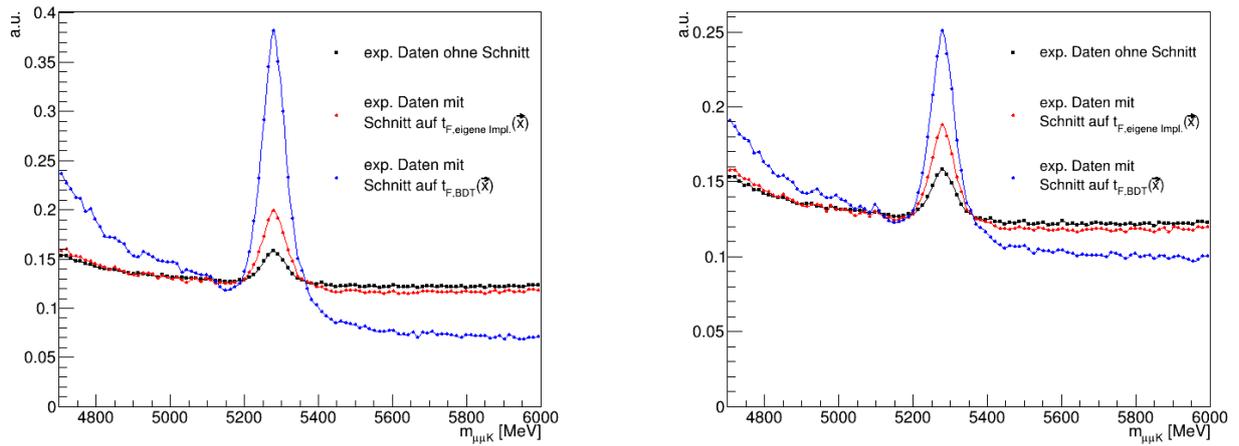


Abbildung 41: Vergleich der selbst implementierten Fisher-Diskriminante mit BDT bei Anwendung auf ATLAS-Daten von 2015 und 2016 entsprechend $\varepsilon_{sig} = 0,8$.

9 Literaturverzeichnis

- [1] CERN Annual report 2021, CERN Geneva, 2022
- [2] Murray Gell-Mann, “A Schematic Model of Baryons and Mesons.”, Phys. Letters 8, 214–215, 1964
- [3] Lyndon Evans, Philip Bryant, JINST 3 S08001, 2008
- [4] The CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, Phys. Letters B. 716, S. 30–61, 2012
- [5] The ATLAS Collaboration, “Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC”, Phys. Letters B. 716, S. 1–29, 2012
- [6] Glen Cowan, “Statistical Data Analysis”, Oxford University Press, 1998
- [7] Andreas Höcker et al., “TMVA - Toolkit for Multivariate Data Analysis”, CERN-OPEN-2007-007, 2007
- [8] Rene Brun, Fons Rademakers, “ROOT: An object oriented data analysis framework”, Nucl. Instrum. Meth. A, Nr. 389, S. 81-86, 1997
- [9] The ATLAS Collaboration, JINST 3 S08003, 2008
- [10] The CERN accelerator complex, CERN Geneva, 2019
- [11] Ronald Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01, 2022
- [12] Gernot Münster, “Von der Quantenfeldtheorie zum Standardmodell”, Neue Auflage, de Gruyter, 2019
- [13] David Galbraith, Charsten Burgard, “Diagram of the Standard Model”, CERN Webfest, 2012
- [14] Makoto Kobayashi, Toshihide Maskawa, Prog. Theor. Phys. 49 652, 1973
- [15] Cheng Li, Bingyu Wang, “Fisher Linear Discriminant Analysis”, German University in Cairo, 2014
- [16] Scott Dean Metzler, “Measurement of the ratio of branching ratios ($B^+ \rightarrow J/\psi\pi^+$)/($B^+ \rightarrow J/\psi^+$) and search for B_C^+ meson”, PhD thesis, University of Pennsylvania, 1996
- [17] Bakul Gaur, “Measurement of the Reference Channel $B^\pm \rightarrow J/\psi K^\pm$ for the Rare Decay $B_S^0 \rightarrow \mu^+\mu^-$ with the ATLAS Experiment”, PhD thesis, University of Siegen, 2013

10 Abkürzungsverzeichnis

- CERN: Conseil Européen pour la Recherche Nucléaire
- CMS: Compact Muon Solenoid
- LHCb: Large Hadron Collider beauty
- ATLAS: A Toroidal LHC ApparatuS
- ALICE: A Large Ion Collider Experiment
- ROC: Receiver Operator Characteristic
- SM: Standard Modell
- AUC: Area under Curve
- TMVA: Toolkit for Multivariate Data Analysis
- BDT: Boosted Decision Tree
- MC: Monte Carlo
- QCD: Quantenchromodynamik
- CKM: Cabibbo-Kobayashi-Masakawa
- QED: Quantenelektrodynamik
- MDT: Monitored Drift Tube
- HCAL: Hadronisches Kalorimeter (Hadronic Calorimeter)
- ECAL: Elektronisches Kalorimeter (Electromagnetic Calorimeter)
- ROOT: am CERN entwickeltes object oriented Toolkit zur Datenanalyse
- TRT: Transition Radiation Tracker
- SCT: Semiconductor Tracker
- LHC: Large Hadron Collider